

Capacity-Achieving Iterative LMMSE Detection for MIMO-NOMA Systems

Lei Liu, *Student Member, IEEE*, Chau Yuen, *Senior Member, IEEE*,
Yong Liang Guan, *Senior Member, IEEE*, and Ying Li, *Member, IEEE*

Abstract

This paper considers a *Iterative Linear Minimum Mean Square Error* (LMMSE) detection for the uplink *Multiuser Multiple-Input and Multiple-Output* (MU-MIMO) systems with *Non-Orthogonal Multiple Access* (NOMA), in which all the users interfere with each other both in the time domain and frequency domain. It is well known that the Iterative LMMSE detection greatly reduces the system computational complexity by departing the overall processing into many low-complexity distributed calculations that can be executed in parallel. However, it is generally considered to be suboptimal and achieves relatively poor performance due to its suboptimal detector. In this paper, we firstly present the matching conditions and area theorems for the iterative detection of the MIMO-NOMA systems. Based on the matching conditions and area theorems, the achievable rate region of the Iterative LMMSE detection is analysed. Interestingly, we prove that by properly design the Iterative LMMSE detection, it can achieve (i) the optimal capacity of symmetric MIMO-NOMA system, (ii) the optimal sum capacity of asymmetric MIMO-NOMA system, (iii) all the maximal extreme points in the capacity region of asymmetric MIMO-NOMA system, (iv) the whole capacity region of two-user and three-user asymmetric MIMO-NOMA systems, in a distributed manner for all cases. Finally, a practical Iterative LMMSE detection design is also proposed for the general asymmetric MIMO-NOMA systems.

Index Terms

MU-MIMO, Non-Orthogonal Multiple Access, Iterative LMMSE detection, Low-complexity, Superposition coded modulation, Achievable rate, Capacity region achieving.

Lei Liu and Ying Li are with the State Key Lab of Integrated Services Networks, Xidian University, Xian, 710071, China (e-mail: lliu_0@stu.xidian.edu.cn, yli@mail.xidian.edu.cn).

Chau Yuen is with the Singapore University of Technology and Design, Singapore (e-mail: yuenchau@sutd.edu.sg).

Yong Liang Guan is with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore (e-mail: eylguan@ntu.edu.sg).

This paper was presented in part at 2016 IEEE International Conference on Communications.

I. INTRODUCTION

Non-Orthogonal Multiple Access (NOMA) has attracted a lot of attentions and been recognized one of the key radio access technologies for the fifth generation (5G) mobile networks [1]–[10]. The key concept behind NOMA is that all the users are allowed to be superimposed at the receiver in the same time/code/frequency domain to significantly increase the spectral efficiency and reduce latency in the 5G communications systems [5]–[9]. In addition, NOMA can be combined with the *Multiuser Multiple-Input and Multiple-Output* (MU-MIMO) [11]–[14], which is a key technology for wireless communication standards like IEEE 802.11 (Wi-Fi), WiMAX and Long Term Evolution (4G), to further enhance the system performance [10]. The MU-MIMO with NOMA is the well known MIMO-NOMA that is discussed in this paper.

Unlike the *Orthogonal Multiple Access* (OMA) systems, *e.g.* the *Time Division Multiple Access* (TDMA) and *Orthogonal Frequency Multiple Access* (OFMA) *ect.*, the signal processing in the NOMA systems will cost higher complexity and higher energy consumption at the *base station* (BS) [12], [13]. Low-complexity uplink detection for MIMO-NOMA is a challenging problem due to the non-orthogonal interference between the users [3]–[5], [12], especially when the number of users and the number of antennas are large. The optimal *multiuser detector* (MUD) for the MIMO-NOMA, such as the *maximum a posterior probability* (MAP) detector or *maximum likelihood* (ML) detector, was proven to be an NP-hard and non-deterministic polynomial-time complete (NP-complete) problem [15], [16]. Furthermore, the complexity of optimal MUD grows exponentially with the number of users or the number of antennas at the BS, and polynomially with the size of the signal constellation [17]–[19]. Many low-complexity linear detections such as *Matched Filter* (MF), *Zero-Forcing* (ZF) receiver [20], *Minimum Mean Square Error* (MMSE) detector and *Message Passing Detector* (MPD) [21]–[23] are proposed for the practical systems. In addition, some iterative methods such as *Jacobi method*, *Richardson method* [24]–[26], *Belief Propagation* (BP) method, and iterative MPD [27]–[30] are put forward to further reduce the computational complexity by avoiding the unfavorable matrix inversion in the linear detections. Although these detectors are attractive from the complexity view point, they achieve relatively poor performance and are considered to be sub-optimal MUDs for the MIMO-NOMA systems.

Successive Interference Cancellation (SIC) based on the low-complexity detections is one of the key technology to improve the system performance. It is well known that for the *multiuser*

access channel (MAC), the SIC is an optimal strategy and can achieve the whole multiuser capacity region with time-sharing technology [31], [32]. In [33], [34], it is showed that the ZF-SIC detector is sub-optimal and can not achieve the full diversity of the MIMO system. In addition, the MMSE-SIC detector [35]–[37] has been proposed to achieve optimal performance [20]. However, the SIC decoding still has some disadvantages when applying to the practical MIMO-NOMA systems, because it needs to preset the decoding order for all the users and assumes that all the previous users' messages are recovered correctly before we decode the next user's message. Furthermore, the decoding order changes with the different channel state and different *Quality of Service* (QoS). These greatly increase the time delay, introduce error propagation during the decoding process, and require additional spectrum resource for the users and base station to achieve the preseted decoding order. To achieve the whole capacity region of the MIMO-NOMA systems, time-sharing should be used, which needs cooperation between the users. In addition, the complexity of SIC decoding is also too high to apply to the real system when the number of antennas at the BS is large [12], [20], [33].

The efficient iterative detection that exchanges soft information of the low-complexity detector with the user decoders is widely used for the practical MIMO-NOMA systems [26]–[28]. This is a fundamental technology for the non-orthogonal MAC like the *Code Division Multiple Access* (CDMA) systems [20], [33] and the *Interleave Division Multiple Access* (IDMA) systems [38]–[40]. Various iterative detectors, such as Iterative *Linear MMSE* (LMMSE) detector, iterative BP detector and iterative MPD, are proposed to achieve a good performance [41]–[45]. The iterative detection is a low-complexity parallel joint decoding method, in which the time delay and error propagation can be greatly reduced. In addition, no user cooperation and additional spectrum resource will be wasted on the preseted decoding order in this method, and the complexity becomes much lower at the same time as the overall receiver is departed into many parallel distributed processors. However, in general, the joint iterative detection structure cannot achieve the perfect performance and is considered to be sub-optimal [33]. Therefore, the achievable rate region analysis of the MIMO-NOMA systems with iterative detection is an intriguing problem.

The *Extrinsic Information Transfer* (EXIT) [48], [49], *MSE-based Transfer Chart* (MSTC) [45]–[47], area theorem and matching theorem [45]–[49] are the main analysis methods of the system achievable rate or the system *Bit Error Rate* (BER) performance. It is proved that a well-designed single-code with linear precoding and Iterative LMMSE detection achieves the

capacity of the MIMO systems [45]. However, this results is based on the *point-to-point* (P2P) MIMO systems. For the multi-user MIMO-NOMA systems, this problem becomes much more complicated, because the whole achievable rate region that contains the achievable rates of all the users remains unknown. In addition, in the multi-user MIMO-NOMA systems, the users interfere with each other and their decoding processes interact each other based on the detector, which results in a much more complicated MST functions and area theorems. As a result, it is more difficult to analyse the achievable rates of all users for the MIMO-NOMA systems.

In this paper, we consider the uplink MIMO-NOMA communication network with a low-complexity Iterative LMMSE detection, where the multiple users communicate with the BS using the same time, frequency, and spreading code resources. The achievable rate region analysis of the Iterative LMMSE detection is provided, which shows that Iterative LMMSE can be rate region optimal for the MIMO-NOMA systems if properly designed. The contributions of this paper are summarized as follows.

- a) For the MIMO-NOMA systems, matching conditions of the iterative detection are proposed. Based on these matching conditions, area theorems of MIMO-NOMA systems with iterative detection are proposed.
- b) With the proposed matching conditions and area theorems, the design and achievable rate analysis of Iterative LMMSE detection for the MIMO-NOMA systems are provided.
- c) We prove that the properly designed Iterative LMMSE detection (i) is capacity achieving for symmetric MIMO-NOMA systems, (ii) is sum capacity achieving for the asymmetric MIMO-NOMA systems, (iii) achieves all the maximal extreme points in the capacity region of asymmetric MIMO-NOMA system, and (iv) achieves the whole capacity region of two-user and three-user asymmetric MIMO-NOMA systems.
- d) For the general asymmetric MIMO-NOMA systems, a numerical algorithm for the practical Iterative LMMSE detection design is provided.

This paper is organized as follows. In Section II, the MIMO-NOMA system model and iterative detection are introduced. The matching conditions and area theorems for the MIMO-NOMA systems are elaborated in Section III. Section IV provides the achievable rate region analysis for the MIMO-NOMA systems with Iterative LMMSE detection. Some special cases are shown in Section V, and we end with conclusions in Section VI.

II. SYSTEM MODEL AND ITERATIVE DETECTION

In this section, the MIMO-NOMA system model and some preliminaries about the iterative detection for the MIMO-NOMA systems are introduced.

A. System Model

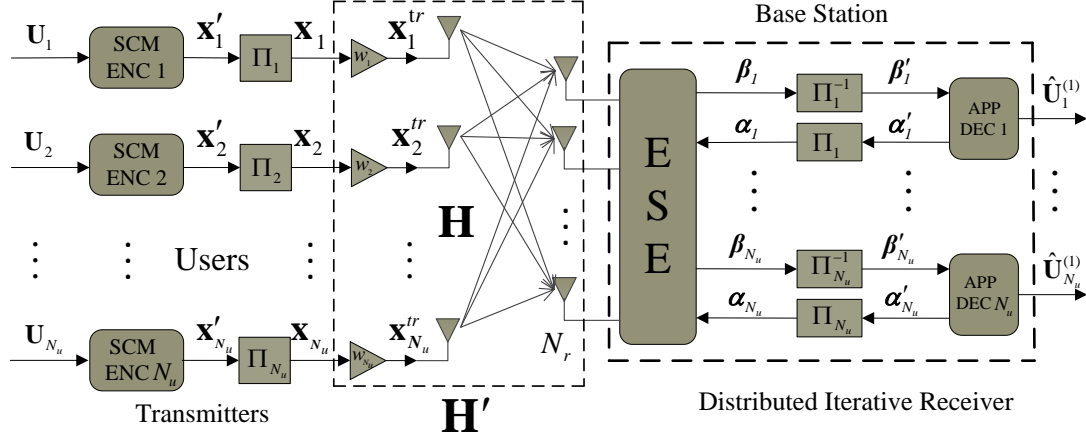


Fig. 1. Block diagram of the MIMO-NOMA system. SCM ENC is the Superposition Coded Modulation Encoder and APP DEC is the A Posteriori Probability Decoder. Π_i denotes the Interleaver and Π_i^{-1} denotes the De-Interleaver. ESE represents the Elementary Signal Estimator. \mathbf{H}' contains the small-scale fading channel \mathbf{H} , and w_i^2 , $i \in \mathcal{N}_u$ denote the power constraint parameters or the large scale fading of the users.

Consider a uplink MU-MIMO system as showed in Fig 1. In this system, N_u autonomous single-antenna terminals simultaneously communicate with an array of N_r antennas of the BS [12], [14] at the same time and under the same frequency. Here, N_u and N_r can be any finite positive integers. Unlike the TDMA systems that the user are orthogonal in the time domain or the OFMA systems that the user are orthogonal in the frequency domain, all the users interfere with each other at the receiver and are non-orthogonal both in the time domain and frequency domain in the NOMA systems. This is the reason why we call it MIMO-NOMA. Then, the $N_r \times 1$ received signal vector $\mathbf{y}(t)$ at time t at the BS can be represented as

$$\mathbf{y}(t) = \mathbf{H}\mathbf{x}^{tr}(t) + \mathbf{n}(t), \quad t \in \mathcal{N}, \quad (1)$$

where $\mathcal{N} = \{1, \dots, N\}$, \mathbf{H} is a $N_r \times N_u$ channel matrix, $\mathbf{n}(t) \sim \mathcal{CN}^{N_r}(0, \sigma_n^2)$ is the $N_r \times 1$ independent additive white Gaussian noise (AWGN) vector at time t , $\mathbf{x}^{tr}(t) = [x_1^{tr}(t), \dots, x_{N_u}^{tr}(t)]^T$

is the message vector sent from N_u users and $\mathbf{y}(t)$ is the message vector got at the N_r receive antennas at time t . In this paper, we consider the fading channels, which means \mathbf{H} denotes the Rayleigh fading channel matrix whose entries are *independently and identically distributed* (i.i.d.) with normal distribution $\mathcal{CN}(0, 1)$, but it is fixed during the transmission. The BS tries to estimate the sources from the received signal vector $\mathbf{y}(t)$, $t \in \mathcal{N}$. Note that the channel matrix \mathbf{H} can be usually achieved by time-domain and/or frequency-domain training pilots. In this paper, we assume that the *Channel State Information* (CSI) \mathbf{H} is known at the BS .

B. Transmitters

As illustrated in Fig. 1, at user i , an information sequence \mathbf{U}_i is encoded by a channel code with rate R_i into a N -length coded sequence \mathbf{x}'_i , $i \in \mathcal{N}_u$, $\mathcal{N}_u = \{1, 2, \dots, N_u\}$ and then interleaved by an N -length independent random interleaver Π_i and get $\mathbf{x}_i = [x_{i,1}, x_{i,2}, \dots, x_{i,N}]^T$. The interleavers here is used for improving the system performance by enhancing the randomness of the messages or the channel noise and avoiding the short cycles in the system factor graph [38], [39], [50]. We assume that each $x_{i,t}$ is randomly and uniformly taken over the points in a discrete signaling constellation $\mathcal{S} = \{s_1, s_2, \dots, s_{|\mathcal{S}|}\}$. This assumption does not lose any generality since signal shaping (e.g., to approach Gaussian signaling) can be realized by properly designing the constellation points of \mathcal{S} . After that, the \mathbf{x}_i is scaled with w_i , which denotes the power constraint or the large-scale fading coefficient of each user, and we then get the transmitting \mathbf{x}_i^{tr} , $i \in \mathcal{N}_u$. Let $K_{\mathbf{x}}$ be power constraint diagonal matrix whose diagonal elements are w_i^2 , $i \in \mathcal{N}_u$. Therefore, the system model can be rewritten as

$$\begin{aligned} \mathbf{y}_t &= \mathbf{H} K_{\mathbf{x}}^{1/2} \mathbf{x}(t) + \mathbf{n}(t) \\ &= \mathbf{H}' \mathbf{x}(t) + \mathbf{n}(t), \quad t \in \mathcal{N}, \end{aligned} \quad (2)$$

where $K_{\mathbf{x}}^{1/2} = \text{diag}\{w_1, \dots, w_{N_u}\}$, $\mathbf{H}' = \mathbf{H} K_{\mathbf{x}}^{1/2}$ is an equivalent channel matrix.

C. Iterative Receiver

At the base station, the received signals $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_N]$ and message α_i from the decoder are sent to a low-complexity elementary signal estimator (ESE) to estimate the extrinsic message β_i , which is then deinterleaved with Π_i^{-1} into β'_i . The corresponding single-user decoder employs β'_i as the prior message to calculate the extrinsic message α'_i . Similarly, this extrinsic message

is interleaved by Π_i to obtain the prior information α_i for the ESE. Repeat this process until the maximum number of iteration is achieved or the messages are recovered correctly. The messages α_i , α'_i , β_i and β'_i , $i \in \mathcal{N}_u$, are defined as follows.

Let α_i be an N -length message vector as

$$\alpha_i = [\alpha_{i,1}, \alpha_{i,2}, \dots, \alpha_{i,N}]^T \quad (3)$$

and $\alpha_{i,t}$ is a vector containing the likelihood values of $x_{i,t}$, i.e.,

$$\alpha_{i,t} = \{\alpha_{i,t}(1), \alpha_{i,t}(2), \dots, \alpha_{i,t}(|\mathcal{S}|)\}, \quad (4)$$

where $\alpha_{i,t}(k)$ denotes the prior likelihood of $x_{i,t} = s_k \in \mathcal{S}$ for the detector, and

$$\sum_{k=1}^{|\mathcal{S}|} \alpha_{i,t}(k) = 1. \quad (5)$$

Similar definitions applies to β'_i , β_i and α'_i . The auxiliary random variables $a_{i,t}$ and $b'_{i,t}$ denote the information carried by $\alpha_{i,t}$ and $\beta'_{i,t}$ respectively, which are defined as

$$p(a_{i,t}|x_i(t) = s_k) = \alpha_{i,t}(k) \quad \text{and} \quad p(b'_{i,t}|x'_i(t) = s_k) = \beta'_{i,t}(k). \quad (6)$$

Let $\mathbf{a}_i = [a_{i,1}, \dots, a_{i,N_u}]$ and $\mathbf{b}'_i = [b'_{i,1}, \dots, b'_{i,N_u}]$, for $i \in \mathcal{N}$. To simplify to analysis of the iterative process, the following two assumptions are made for the ESE and decoders.

Assumption 1: For the estimator, each $x_i(t)$ is independently chosen from \mathcal{S} with equal probability, i.e., $p(x_i(t) = s_k) = 1/|\mathcal{S}|$, for any i , k and t ; the messages \mathbf{a}_i , $i \in \mathcal{N}_u$ are independent with each other, and the entries of \mathbf{a}_i are i.i.d. given \mathbf{x}_i , i.e.,

$$p(\mathbf{a}_i|\mathbf{x}_i) = \prod_{t=1}^N p(a_{i,t}|x_{i,t}) \quad (7)$$

and

$$p(a_{i,t_1} = s_l | x_{i,t_1} = s_k) = p(a_{i,t_2} = s_l | x_{i,t_2} = s_k) \quad (8)$$

for any $i \in \mathcal{N}_u$, $s_k, s_l \in \mathcal{S}$ and $t_1, t_2 \in \mathcal{N}$.

Assumption 2: For the decoder, the messages \mathbf{b}'_i , $i \in \mathcal{N}_u$ are independent with each other, and the entries of \mathbf{b}'_i are i.i.d. given \mathbf{x}'_i , i.e.,

$$p(\mathbf{b}'_i|\mathbf{x}'_i) = \prod_{t=1}^N p(b'_{i,t}|x'_{i,t}) \quad (9)$$

and

$$p(b'_{i,t_1} = s_l | x'_{i,t_1} = s_k) = p(b'_{i,t_2} = s_l | x'_{i,t_2} = s_k) \quad (10)$$

for any $i \in \mathcal{N}_u$, $s_k, s_l \in \mathcal{S}$ and $t_1, t_2 \in \mathcal{N}$.

Assumptions 1 and 2 decompose the overall process into the local processors such as the estimator and decoders, which simplifies the analysis of the iterative process. These assumptions are widely used in iterative decoding and turbo equalization algorithms. Actually, the messages $\beta'_i, \beta_i, \alpha_i$ and α'_i can be replaced by the auxiliary random variables $\mathbf{b}'_i, \mathbf{b}_i, \mathbf{a}_i$ and \mathbf{a}'_i respectively.

1) *Element Signal Estimator (ESE)*: The extrinsic messages of the ESE is defined as

$$\beta_{i,t}(k) = p(x_{i,t} = s_k | \alpha_{\sim i,t}, \mathbf{y}_t), \text{ for } i \in \mathcal{N}_u \text{ and } t \in \mathcal{N}, \quad (11)$$

where $\alpha_{\sim i,t}$ represents the vector obtained by deleting the i th entry of $[\alpha_{1,t}, \dots, \alpha_{N_u,t}]$. Here, “extrinsic” and “elementary signal estimator” means that the output estimation for $x_{i,t}$, i.e., the t th transmission at user i , is based on the t th receiver \mathbf{y}_t and the messages of the t th symbols from the decoders except user i .

For the MIMO-NOMA systems, it is too complicated to calculate (11) directly. Element signal estimator is an alternative low complexity estimator. For the ESE, with the input prior messages $\alpha_{i,t}$, we can calculate the prior mean and variance of $x_i(t)$, respectively, by

$$\bar{x}_{i,t} = E[x_{i,t} | \alpha_{i,t}] = \sum_{k=1}^{|S|} \alpha_{i,t}(k) s_k, \quad (12)$$

and

$$v_i = E[|x_{i,t} - \bar{x}_{i,t}|^2] = \sum_{k=1}^{|S|} \alpha_{i,t}(k) |s_k - \bar{x}_{i,t}|^2, \quad (13)$$

for any index t . From Assumption 1, v_i is invariant with respect to the index t . Then, the input messages $\alpha_{i,t}$ can be replaced by the $\bar{x}_{i,t}$ and v_i if the messages are Gaussian distributed.

Let $\hat{x}_{i,t}$ be the total estimation of $x_{i,t}$ at the detector, and $v_{\hat{x}_i}$ be the corresponding estimated deviation. The extrinsic mean and variance for $x_{i,t}$ (denoted by u_i and $b_{i,t}$) by excluding the contribution of the prior message α_i according to the message combining rule [21]–[23] as

$$u_i^{-1} = v_{\hat{x}_i}^{-1} - v_i^{-1} \quad (14)$$

and

$$\frac{b_{i,t}}{u_i} = \frac{\hat{x}_{i,t}}{v_{\hat{x}_i}} - \frac{\bar{x}_{i,t}}{v_i}. \quad (15)$$

Finally, the extrinsic message of the estimator can be given by

$$\beta_{i,t}(k) = \frac{p(b_{i,t}|x_{i,t} = s_k)}{\sum_{k=1}^{|S|} p(b_{i,t}|x_{i,t} = s_k)}. \quad (16)$$

The overall complexity has been reduced by the local processing of the ESE and distributed decoders. This is very important for the application for the practical systems. We will show that not only the complexity of the iterative receiver are greatly reduced, but also it is sum capacity achieving if the ESE and channel codes of each user are properly designed.

2) *A Posteriori Probability (APP) Decoding*: The decoder employees the APP decoding based on the input message β_i , $i \in \mathcal{N}_u$. The output extrinsic message of the decoder is defined as

$$\alpha'_{i,t}(k) = p(x_i(t) = s_k | \beta'_{\sim i}), \quad (17)$$

where $i \in \mathcal{N}_u$, $t \in \mathcal{N}$ and $k \in \mathcal{S}$.

Assumption 3: The local decoder is APP decoding.

Although computational complexity of the APP decoding is too high to apply in practical systems, low-complexity message-passing algorithms can be used to achieve near-optimal performance [51]. Assumption 3 is introduced to simplify our analysis. As the message passing decoding is well-studied in the literature, and thus the details are omitted here.

D. Capacity region of the MIMO-NOMA systems

Let \mathbf{Y} denote the received random variable vector at the BS, and \mathbf{X} represent the received random variable vector from the users. Assume $S \subseteq \mathcal{N}_u$, $S^c \subseteq \mathcal{N}_u/S$ and $S \cup S^c = \mathcal{N}_u$, the partial channel matrix is denoted as $\mathbf{H}'_S = [\mathbf{h}'_i]_{N_r \times |S|}$, where \mathbf{h}'_i is the i th column of \mathbf{H}' and $i \in S$. Similar definition applies to the partial random vector \mathbf{X}_S and the partial matrix $K_{\mathbf{x}_S}$. The rate of user i is R_i and $R_S = \sum_{i \in S} R_i$ represents the sum rate of the users in set S . Then, for our MIMO-NOMA systems (2), the capacity region \mathcal{R}_S is given as follows [31], [32].

$$\begin{aligned} R_S &\leq I(\mathbf{Y}; \mathbf{X}_S | \mathbf{X}_{S^c}) \\ &= \log \det \left(\mathbf{I}_{|S|} + \frac{1}{\sigma_n^2} \mathbf{H}'_S{}^H \mathbf{H}'_S \right), \text{ for } \forall S \subseteq \mathcal{N}_u \text{ and } S^c \subseteq \mathcal{N}_u/S. \end{aligned} \quad (18)$$

Therefore, the sum rate of the MU-MIMO systems is denoted as

$$\begin{aligned} R_{sum} &= R_{\mathcal{N}_u} \leq I(\mathbf{Y}; \mathbf{X}_{\mathcal{N}_u}) \\ &= \log \det \left(\mathbf{I}_{N_u} + \frac{1}{\sigma_n^2} \mathbf{H}'^H \mathbf{H}' \right). \end{aligned} \quad (19)$$

Capacity Region Domination Lemma: [52] All point in the capacity region \mathcal{R}_S is dominated by a convex combination of the following $(N_u!)$ maximal extreme points.

$$\begin{cases} R_{k_1} = I(\mathbf{Y}; \mathbf{X}_{S_1}) = \log \det \left(\mathbf{I}_{N_u} + \frac{1}{\sigma_n^2} \mathbf{H}'^H \mathbf{H}' \right) - \log \det \left(\mathbf{I}_{|S_1^c|} + \frac{1}{\sigma_n^2} \mathbf{H}'_{S_1^c}{}^H \mathbf{H}'_{S_1^c} \right), \\ \vdots \\ R_{k_{N_u-1}} = I(\mathbf{Y}; \mathbf{X}_{k_{N_u-1}} | \mathbf{X}_{S_{N_u-2}}) = \log \det \left(\mathbf{I}_{|S_{N_u-2}^c|} + \frac{1}{\sigma_n^2} \mathbf{H}'_{S_{N_u-2}^c}{}^H \mathbf{H}'_{S_{N_u-2}^c} \right) - \log \left(1 + \frac{1}{\sigma_n^2} \mathbf{h}'_{k_{N_u}}{}^H \mathbf{h}'_{k_{N_u}} \right), \\ R_{k_{N_u}} = I(\mathbf{Y}; \mathbf{X}_{k_{N_u}} | \mathbf{X}_{S_{N_u-1}}) = \log \det \left(\mathbf{I}_{|S_{N_u-1}^c|} + \frac{1}{\sigma_n^2} \mathbf{H}'_{S_{N_u-1}^c}{}^H \mathbf{H}'_{S_{N_u-1}^c} \right) = \log \left(1 + \frac{1}{\sigma_n^2} \mathbf{h}'_{k_{N_u}}{}^H \mathbf{h}'_{k_{N_u}} \right), \end{cases} \quad (20)$$

where (k_1, \dots, k_{N_u}) is a permutation of $(1, 2, \dots, N_u)$, $S_i = \{k_1, \dots, k_i\}$ and $S_i^c = \mathcal{N}_u / \{k_1, \dots, k_i\}$ for $i = 1, \dots, N_u - 1$.

III. MATCHING CONDITIONS AND AREA THEOREMS FOR MIMO-NOMA SYSTEMS

In this section, we proposed the matching conditions and area properties of the iterative MIMO-NOMA systems from their *SINR-Variance* transfer functions. In [45]–[47], the *I-MMSE* theorem and the area theorems for the P2P communication systems are proposed for its achievable rate analysis. In this section, these results are generalized to the MIMO-NOMA systems, which is much more complicated than the P2P case as all the users' signals interfere with each other and their decoding processes interact each other based on the detector. We will see that the achievable rate of each user in the iterative MIMO-NOMA systems can be derived based on these useful theorems and techniques.

A. Characterization of ESE

We first define the *SINR-Variance* transfer function of user i of element signal estimator as

$$\phi_i(\mathbf{v}_{\bar{\mathbf{x}}}) = v_{\hat{x}_i}^{-1} - v_i^{-1}, \quad \text{for } i \in \mathcal{N}_u, \quad (21)$$

where $\mathbf{v}_{\bar{\mathbf{x}}} = [v_1, \dots, v_{N_u}]$, $v_{\hat{x}_i}$ is the i th diagonal element of covariance matrix $\mathbf{V}_{\hat{\mathbf{x}}}$, and v_i is the input variance of user i , i.e., the i th diagonal element of $\mathbf{V}_{\bar{\mathbf{x}}}$. Actually, $\phi_i(\mathbf{v})$ denotes the output extrinsic *SINR* of user i at the estimator. Similarly, the total MSE of user i at the estimator is

$$\text{mmse}_{\text{tot},i}^{\text{ese}}(\mathbf{v}_{\bar{\mathbf{x}}}) = v_{\hat{x}_i}. \quad (22)$$

It should be noted that the variance v_i varies from 0 to 1, because the signal power is normalized to 1. From (19), the output estimation of user i depends on the input variances

of all the users. Thus, the *SINR-Variance* transfer functions of all users interact with each other. In addition, $\phi_i(\mathbf{v}_{\bar{\mathbf{x}}})$ is monotonically decreasing in \mathbf{v} , which means the lower input variances of the users, the higher the output *SINR* of the estimator. The next Gaussian assumption is used to simplify the system analysis, which is a common assumption in many works [53], [54].

Assumption 4: Let $\boldsymbol{\rho} = [\rho_1, \dots, \rho_{N_u}]$, $\boldsymbol{\phi}(\mathbf{v}_{\bar{\mathbf{x}}}) = [\phi_1(\mathbf{v}_{\bar{\mathbf{x}}}), \dots, \phi_{N_u}(\mathbf{v}_{\bar{\mathbf{x}}})]$. The outputs of the estimator can be approximated as the observations from AWGN channels and the related SINR is denoted by $\boldsymbol{\rho} = \boldsymbol{\phi}(\mathbf{v}_{\bar{\mathbf{x}}})$.

B. Characterization of APP Decoder

We now consider the *SINR-Variance* transfer functions of the decoders. From Assumption 4, the input of the each decoder \mathbf{b}'_i are equivalent as the independent observations over an AWGN channel with $\text{SNR}_i = \rho_i$. As the LMMSE estimator only depends on the variance of input messages, we thus use the *SINR-Variance* transfer functions to describe the decoders. The output extrinsic variance of APP decoder is defined as

$$v'_i = v'_{i,t} = \text{MMSE}(x'_{i,t} | \mathbf{b}'_{i,\sim t}) = E \left[|x'_{i,t} - E[x'_{i,t} | \mathbf{b}'_{i,\sim t}]|^2 \right], \quad (23)$$

where the input messages β'_i are replaced by auxiliary random variable \mathbf{b}'_i , and the expectation is taken over the joint probability space of \mathbf{x}'_i and \mathbf{b}'_i . The first equation is derived from the i.i.d. Assumption 2. Therefore, we define the *SINR-Variance* transfer function of the decoders as

$$v_i = \psi_i(\rho_i) \quad (24)$$

for any $i \in \mathcal{N}_u$, where ψ_i can be different transfer function with different i . Let $\boldsymbol{\psi}(\boldsymbol{\rho}) = [\psi_1(\rho_1), \dots, \psi_{N_u}(\rho_{N_u})]$, and we have

$$\mathbf{v}_{\bar{\mathbf{x}}} = \boldsymbol{\psi}(\boldsymbol{\rho}). \quad (25)$$

C. SINR-Variance Transfer Chart and Matching Conditions

The LMMSE estimator is described by $\boldsymbol{\rho} = \boldsymbol{\phi}(\mathbf{v}_{\bar{\mathbf{x}}})$, and the decoders can be described by $\mathbf{v}_{\bar{\mathbf{x}}} = \boldsymbol{\psi}(\boldsymbol{\rho})$. Therefore, the iterative detection performs iteration between the estimator and the decoders and can be tracked by $\boldsymbol{\rho}$ and $\mathbf{v}_{\bar{\mathbf{x}}}$. Let τ represent the τ -th iteration. We then have

$$\boldsymbol{\rho}(\tau) = \boldsymbol{\phi}(\mathbf{v}_{\bar{\mathbf{x}}}(\tau - 1)) \text{ and } \mathbf{v}_{\bar{\mathbf{x}}}(\tau) = \boldsymbol{\psi}(\boldsymbol{\rho}(\tau)), \quad \tau = 1, 2, \dots \quad (26)$$

The iteration converges to a fix point $\mathbf{v}_{\bar{\mathbf{x}}}^*$ which satisfies

$$\phi(\mathbf{v}_{\bar{\mathbf{x}}}^*) = \psi^{-1}(\mathbf{v}_{\bar{\mathbf{x}}}^*) \text{ and } \phi(\mathbf{v}_{\bar{\mathbf{x}}}) > \psi^{-1}(\mathbf{v}_{\bar{\mathbf{x}}}), \quad \text{for } \mathbf{v}_{\bar{\mathbf{x}}}^* < \mathbf{v}_{\bar{\mathbf{x}}} \leq \mathbf{1}, \quad (27)$$

where $\psi^{-1}(\cdot)$ denotes the inverse of $\psi(\cdot)$ which exists since $\psi(\cdot)$ is continuous and monotonic [55]. It should be mentioned that the equations and inequalities for the vectors or matrixes in this paper correspond to the component-wise inequality. The inequality $\mathbf{v}_{\bar{\mathbf{x}}} \leq \mathbf{1}$ comes from the normalized signal powers of $\mathbf{x}(t)$, $t \in \mathcal{N}$. If $\mathbf{v}_{\bar{\mathbf{x}}}^* = \mathbf{0}$, then all the transmissions can be recovered correctly, which means that $\phi(\mathbf{v}_{\bar{\mathbf{x}}}) > \psi^{-1}(\mathbf{v}_{\bar{\mathbf{x}}})$ for any available $\mathbf{v}_{\bar{\mathbf{x}}}$, i.e., the inverse transfer function of the decoder $\psi^{-1}(\mathbf{v}_{\bar{\mathbf{x}}})$ lies below the transfer function of the estimator $\phi(\mathbf{v}_{\bar{\mathbf{x}}})$.

The estimator and decoders are matched if

$$\phi(\mathbf{v}_{\bar{\mathbf{x}}}) = \psi^{-1}(\mathbf{v}_{\bar{\mathbf{x}}}), \quad \text{for } \mathbf{0} < \mathbf{v}_{\bar{\mathbf{x}}} \leq \mathbf{1}. \quad (28)$$

It means that $\phi_i(\mathbf{v}_{\bar{\mathbf{x}}}) = \psi_i^{-1}(v_i)$ for any $i \in \mathcal{N}_u$. The matched transfer function not only maximizes the rate of the code, but also ensures the transmitting signals can be perfectly recovered. Note that: $\phi_i(\mathbf{1}) > 0$ as the estimator always use the information from the channel; and $\phi_i(\mathbf{0}) > 1$ as the estimator cannot remove the uncertainty introduced by the channel noise. Therefore, we have the following proposition.

Proposition 1: For any $i \in \mathcal{N}_u$, the matching conditions of the iterative MIMO-NOMA systems can be rewritten as

$$\psi_i(\rho_i) = \phi_i^{-1}(\phi_i(\mathbf{1})) = 1, \quad \text{for } 0 \leq \rho_i < \phi_i(\mathbf{1}); \quad (29)$$

$$\psi_i(\rho_i) = \phi_i^{-1}(\rho_i), \quad \text{for } \phi_i(\mathbf{1}) \leq \rho_i < \phi_i(\mathbf{0}); \quad (30)$$

$$\psi_i(\rho_i) = 0, \quad \text{for } \phi_i(\mathbf{0}) \leq \rho_i < \infty. \quad (31)$$

Similarly, $\phi_i^{-1}(\cdot)$ denotes the inverse of $\phi_i(\cdot)$ corresponding to the variable v_i . These matching conditions are very important for the area theorems in the next section.

D. Area Properties

Let $\text{snr}_{ap,i}^{dec}$ denote the SNR of the prior input messages in decoder i , $\text{snr}_{ext,i}^{ese}$ be the SNR of the output extrinsic messages of user i at the estimator, $\text{mmse}_{tot,i}^{ese}(\cdot)$ represent the total variance of the messages of user i at the LMMSE estimator, and $\text{mmse}_{tot,i}^{dec}(\cdot)$ indicate the total variance of the messages at decoder i . In addition, $\text{snr}_{ext,i}^{ese} = [\text{snr}_{ext,1}^{ese}, \dots, \text{snr}_{ext,N_u}^{ese}]$. Then, we present two

area properties of the estimator and decoders as the following proposition, which will be used to derive the achievable rates of each user.

Proposition 2: The achievable rate R_i of user i and an upper bound of R_i are given as

$$R_i = \int_0^\infty \text{mmse}_{\text{tot},i}^{\text{dec}}(\text{snr}_{\text{ap},i}^{\text{dec}}) d\text{snr}_{\text{ap},i}^{\text{dec}}, \quad (32)$$

$$R_i^{\max} = \int_0^\infty \text{mmse}_{\text{tot},i}^{\text{ese}}(\mathbf{snr}_{\text{ext},i}^{\text{ese}}) d\text{snr}_{\text{ext},i}^{\text{ese}}, \quad (33)$$

and $R_i \leq R_i^{\max}$, $i \in \mathcal{N}_u$, where the equality holds if and only if the SINR-Variance transfer functions of the element signal estimator and decoders for any user are matched with each other, i.e., the matching conditions (28)~(31) hold.

In our MIMO-NOMA system model, from (19) and (24) and with the Gaussian assumptions, we have $\text{snr}_{\text{ap},i}^{\text{dec}} = \rho_i$, $\text{snr}_{\text{ext},i}^{\text{ese}} = \phi_i(\mathbf{v}_{\bar{\mathbf{x}}})$, $\text{mmse}_{\text{tot},i}^{\text{dec}}(\mathbf{snr}_{\text{ap}}^{\text{dec}}) = (\rho_i + \psi_i(\rho_i)^{-1})^{-1}$ and $\text{mmse}_{\text{tot},i}^{\text{ese}}(\mathbf{snr}_{\text{ext},i}^{\text{ese}}) = v_{\hat{x}_i}(\mathbf{v}_{\bar{\mathbf{x}}})$. Therefore, (32) and (33) can be rewritten as following.

Proposition 3: With the SINR-Variance transfer functions $\boldsymbol{\rho} = \boldsymbol{\phi}(\mathbf{v}_{\bar{\mathbf{x}}})$ and $\mathbf{v}_{\bar{\mathbf{x}}} = \boldsymbol{\psi}(\boldsymbol{\rho})$, the achievable rate R_i of user i and an upper bound of R_i are

$$R_i = \int_0^\infty (\rho_i + \psi_i(\rho_i)^{-1})^{-1} d\rho_i, \quad (34)$$

$$R_i^{\max} = \int_0^\infty v_{\hat{x}_i}(\mathbf{v}_{\bar{\mathbf{x}}}) d\phi_i(\mathbf{v}_{\bar{\mathbf{x}}}). \quad (35)$$

and $R_i \leq R_i^{\max}$, $i \in \mathcal{N}_u$, where the equality holds if and only if the SINR-Variance transfer functions of the element signal estimator and decoders for any user are matched with each other i.e., the matching conditions (28)~(31) hold.

Thus, the achievable rates can be calculated directly by (35) or be calculated directly by (34) together with (28) and the matching conditions (29)~(31). From (19), (24), (28) and (35), we can see that all the users' transfer functions interact with each other at the estimator since every output of the estimator depends on the variances of the input messages from all the decoders. In addition, all the users' transfer functions are unknown and need to be properly designed. Thus, it is very hard to calculate the achievable rates directly with (34) and (35). However, for some special cases or with some additional constraints on the users' transfer functions, as we will see in the next section, the achievable rate of each user can then be calculated.

IV. ACHIEVABLE RATE REGION ANALYSIS OF ITERATIVE LMMSE DETECTION

In this section, based on the proposed matching conditions and area properties, the achievable rate of each user is given for the Iterative MIMO-NOMA systems. Here, we use the LMMSE estimator as the ESE estimator and the Superposition Coded Modulation (SCM) codes as the channel codes. As it is very difficult to analyse the achievable rate region for the general case, we first consider the symmetric MIMO-NOMA systems, where the users have the same rate and the same power. We prove that the Iterative LMMSE detection achieves the capacity of symmetric MIMO-NOMA system. This means that for the symmetric MIMO-NOMA systems, the Iterative LMMSE detection is optimal if the channel codes are properly designed. For the general asymmetric MIMO-NOMA systems, we also prove that the sum rate of the Iterative LMMSE detection achieves the sum capacity.

A. LMMSE ESE

LMMSE is an alternative low complexity ESE. In general, LMMSE detection is suboptimal for the multi-user large-scale MIMO-NOMA system with the discrete (or digital) sources [12]. However, for the gaussian sources, LMMSE detection is an optimal linear detector under MSE measure because it minimizes the MSE between sources and estimation [33]. Let $\bar{\mathbf{x}}(t) = [x_{1,t}, \dots, x_{N_u,t}]$ and $\mathbf{V}_{\bar{\mathbf{x}}(t)} = \mathbf{V}_{\bar{\mathbf{x}}} = \text{diag}(v_1, v_2, \dots, v_{N_u})$. The LMMSE detector [20] is given by

$$\begin{aligned}\hat{\mathbf{x}}(t) &= \mathbf{V}_{\hat{\mathbf{x}}}^{-1} [\mathbf{V}_{\bar{\mathbf{x}}}^{-1} \bar{\mathbf{x}}(t) + \sigma_n^{-2} \mathbf{H}'^H \mathbf{y}_t] \\ &= (\sigma_n^{-2} \mathbf{H}'^H \mathbf{H}' + \mathbf{V}_{\bar{\mathbf{x}}}^{-1})^{-1} [\mathbf{V}_{\bar{\mathbf{x}}}^{-1} \bar{\mathbf{x}}(t) + \sigma_n^{-2} \mathbf{H}'^H \mathbf{y}_t],\end{aligned}\quad (36)$$

where $\mathbf{V}_{\hat{\mathbf{x}}} = (\sigma_n^{-2} \mathbf{H}'^H \mathbf{H}' + \mathbf{V}_{\bar{\mathbf{x}}}^{-1})^{-1}$, which denotes the deviation of the estimation to the sources.

Matrix Inversion Lemma: Let \mathbf{A} and \mathbf{D} be positive definite matrixes. For matrixes \mathbf{B} and \mathbf{C} with proper size, we have

$$(\mathbf{A} + \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{D} + \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}\mathbf{A}^{-1}. \quad (37)$$

With the Matrix Inversion Lemma, the LMMSE estimator (36) can be rewritten as

$$\begin{aligned}\hat{\mathbf{x}}(t) &= (\sigma_n^{-2} \mathbf{H}'^H \mathbf{H}' + \mathbf{V}_{\bar{\mathbf{x}}}^{-1})^{-1} [\mathbf{V}_{\bar{\mathbf{x}}}^{-1} \bar{\mathbf{x}}(t) + \sigma_n^{-2} \mathbf{H}'^H \mathbf{y}_t] \\ &= [\mathbf{V}_{\bar{\mathbf{x}}} - \mathbf{V}_{\bar{\mathbf{x}}} \mathbf{H}'^H (\sigma_n^2 \mathbf{I}_{N_r} + \mathbf{H}' \mathbf{V}_{\bar{\mathbf{x}}} \mathbf{H}'^H)^{-1} \mathbf{H}' \mathbf{V}_{\bar{\mathbf{x}}}] [\mathbf{V}_{\bar{\mathbf{x}}}^{-1} \bar{\mathbf{x}}(t) + \sigma_n^{-2} \mathbf{H}'^H \mathbf{y}_t] \\ &= \bar{\mathbf{x}}(t) + \mathbf{V}_{\bar{\mathbf{x}}} \mathbf{H}'^H (\sigma_n^2 \mathbf{I}_{N_r} + \mathbf{H}' \mathbf{V}_{\bar{\mathbf{x}}} \mathbf{H}'^H)^{-1} (\mathbf{y}_t - \mathbf{H}' \bar{\mathbf{x}}(t)).\end{aligned}\quad (38)$$

Therefore, with (15) and (38), we get

$$b_{i,t} = x_{i,t} + n_{i,t}^*, \quad (39)$$

and

$$n_{i,t}^* = \frac{v_i}{v_{\hat{x}_i} \rho_i} v_{\hat{x}_i}^2 \mathbf{h}_i'^H \left(\sigma_n^2 \mathbf{I}_{N_r} + \mathbf{H}' \mathbf{V}_{\bar{\mathbf{x}}} \mathbf{H}'^H \right)^{-1} \left[\mathbf{H}' \left(\mathbf{x}_{\setminus i}(t) - \bar{\mathbf{x}}_{\setminus i}(t) \right) + \mathbf{n}(t) \right]. \quad (40)$$

where $\mathbf{x}_{\setminus i}(t)$ (or $\bar{\mathbf{x}}_{\setminus i}(t)$) denotes the vector whose i th entry of $\mathbf{x}(t)$ (or $\bar{\mathbf{x}}(t)$) is set to zero. Thus, we rewrite the Assumption 4 as follows.

Assumption 5: The equivalent noise $n_{i,t}^$ is independent of $x_{i,t}$ and is Gaussian distributed $n_{i,t} \sim \mathcal{CN}(0, 1/\phi_i(\mathbf{v}_{\bar{\mathbf{x}}}))$, i.e., the output of the LMMSE estimator \mathbf{b}_t is the observation from AWGN channel, i.e., $\mathbf{b}_t = \mathbf{x}(t) + \mathbf{n}_t^*$ with SNRs $\rho = \phi(\mathbf{v}_{\bar{\mathbf{x}}})$.*

Note: Although the overall complexity has been reduced by the local ESE and the distributed decoding processing, the complexity of LMMSE estimator is still very high for the practical applications when the number of users and antennas very large. The high complexity is mainly introduced by the matrix inversion and matrix multiplications in (36). In the practical systems, a lower complexity Gaussian message passing iterative estimator can also be used as the ESE for the iterative receiver of the MIMO-NOMA systems, which was proved that it converges to the LMMSE detector [27].

B. Superposition Coded Modulation APP Decoders

In this subsection, we introduce an important property that is established in [45], which builds the relationship between the rate of the Forward Error Correction (FEC) code and its transfer function $\psi_i(\rho_i)$. The result is based on the SCM [56], [57] and the area theorems [46], [47]. For more details, please refer to [45].

Property of SCM Codes: Assume $\psi(\rho)$ satisfies the following regularity conditions:

- (i) $\psi(0) = 1$ and $\psi(\rho) \geq 0$, for $\rho \in [0, \infty)$;
- (ii) monotonically decreasing in $\rho \in [0, \infty)$;
- (iii) continuous and differentiable in $[0, \infty)$ except for a countable set of values of ρ ;
- (iv) $\lim_{\rho \rightarrow \infty} \rho \psi(\rho) = 0$.

Let Γ_n be an n -layer SCM code with SINR-variance transfer function $\psi^n(\rho)$ and rate R_n . Then, there exist $\{\Gamma_n\}$ such that: (i) $\psi^n(\rho) \leq \psi(\rho)$, for any $\rho \geq 0$ and any integer n ; (ii) as $n \rightarrow \infty$,

$$R_n \rightarrow R(\psi(\rho)). \quad (41)$$

In (41), $R(\psi(\rho))$ denotes the rate of the code whose transfer function is $\psi(\rho)$. It means that there exist such an n -layer SCM code Γ_n whose transfer function can approach the function $\psi(\rho)$ that satisfies the conditions (i)~(iv) with arbitrary small error when n is large enough.

C. Capacity Achieving of Iterative LMMSE Detection for Symmetric Systems

In the subsection, we consider the symmetric MIMO-NOMA system, that is the users have the same power and the same rate constraints, i.e., $K_{\mathbf{x}} = w^2 I$ and $R_i = R_j$, for $i, j \in \mathcal{N}_u$. In addition, we assume that the number of users N_u and the number of antennas N_r are very large (be hundreds or more), i.e., the large-scale MIMO-NOMA case. In this case, since all the users have the same conditions, we thus assume that all the users have the same transfer functions, which means $v_i = v$ and $\rho_i = \rho$, for any $i \in \mathcal{N}_u$. The symmetric and large-scale systems assumptions are used to simplify the analysis of the achievable user rates. Actually, in the next subsection, the achievable rate analysis for the asymmetric systems with finite N_r and N_u shows that this result also works for the practical asymmetric systems.

Based on these assumptions, we have

$$\begin{aligned}
 v_{\hat{x}_i}(\mathbf{v}_{\bar{\mathbf{x}}}) &\approx \frac{1}{N_u} \text{mmse}_{tot}^{ese}(\mathbf{v}_{\bar{\mathbf{x}}}) \\
 &= \frac{1}{N_u} \text{Tr}\{\mathbf{V}_{\bar{\mathbf{x}}}\} \\
 &= \frac{1}{N_u} \text{Tr}\{(\sigma_n^{-2} \mathbf{H}'^H \mathbf{H}' + \mathbf{V}_{\bar{\mathbf{x}}}^{-1})^{-1}\} \\
 &= \frac{1}{N_u} \text{Tr}\{(\sigma_n^{-2} w^2 \mathbf{H}^H \mathbf{H} + v^{-1} \mathbf{I}_{N_u})^{-1}\} \\
 &= v_{\hat{x}}(v),
 \end{aligned} \tag{42}$$

and

$$\begin{aligned}
 \phi_i(\mathbf{v}_{\bar{\mathbf{x}}}) &= v_{\hat{x}_i}^{-1} - v_i^{-1} \\
 &\approx v_{\hat{x}}(v)^{-1} - v^{-1} \\
 &= \frac{1}{N_u} \text{Tr}\{(\sigma_n^{-2} w^2 \mathbf{H}^H \mathbf{H} + v^{-1} \mathbf{I}_{N_u})^{-1}\}^{-1} - v^{-1} \\
 &= \phi(v) \\
 &= \rho.
 \end{aligned} \tag{43}$$

The approximations here is based on the symmetry of the system model and the law of large numbers. Similarly, we have $\psi_i(\rho_i) = \psi(\rho)$, $i \in \mathcal{N}_u$, and the following matching conditions

Proposition 4: For any $i \in \mathcal{N}_u$, the matching conditions of the iterative symmetric MIMO-NOMA system can be rewritten as

$$\psi(\rho) = \phi^{-1}(\phi(1)) = 1, \text{ for } 0 \leq \rho < \phi(1); \quad (44)$$

$$\psi(\rho) = \phi^{-1}(\rho), \text{ for } \phi(1) \leq \rho < \phi(0); \quad (45)$$

$$\psi(\rho) = 0 \text{ for } \phi(0) \leq \rho < \infty. \quad (46)$$

Therefore, the analysis of the transfer functions for the MIMO-NOMA systems are degenerated into that of the single-user and single-antenna systems. Then, we can get the following theorem.

Theorem 1: For a symmetric Large-scale MIMO-NOMA system whose users have: (i) the same rate $R_i = R$ for $i \in \mathcal{N}_u$; (ii) the same power $K_x = w^2 I$, the Iterative LMMSE detection achieves the capacity, i.e., $R_i = \frac{1}{N_u} \log \det \left(I_{N_r} + \frac{w^2}{\sigma_n^2} \mathbf{H} \mathbf{H}^H \right)$ for $i \in \mathcal{N}_u$ and $R_{sum} = \log \det \left(I_{N_r} + \frac{w^2}{\sigma_n^2} \mathbf{H} \mathbf{H}^H \right)$.

Proof: Theorem 1 can be proved either by (34) at the decoder or by (35) at the estimator directly. Here, we show the first proof, and the other one is given in APPENDIX A.

i) User Rate Calculation: From (34), the achievable rate of user i is given by

$$\begin{aligned} R_i &= \int_0^\infty (\rho_i + \psi_i(\rho_i)^{-1})^{-1} d\rho_i \\ &\stackrel{(a)}{\approx} \int_0^\infty (\rho + \psi(\rho)^{-1})^{-1} d\rho \\ &\stackrel{(b)}{\leq} \int_{\phi(1)}^{\phi(0)} \left[\rho + (\phi^{-1}(\rho))^{-1} \right]^{-1} d\rho + \int_0^{\phi(1)} (1 + \rho)^{-1} d\rho \\ &\stackrel{(c)}{=} \int_{v=1}^{v=0} (v^{-1} + \phi(v))^{-1} d\phi(v) + \log(1 + \phi(v)) \\ &\stackrel{(d)}{=} \int_{v=1}^{v=0} v_{\hat{x}}(v) dv_{\hat{x}}(v)^{-1} - \int_{v=1}^{v=0} v_{\hat{x}}(v) dv^{-1} - \log v_{\hat{x}}(v=1) \end{aligned}$$

$$\begin{aligned}
&\stackrel{(e)}{=} - \int_{v=1}^{v=0} \frac{1}{N_u} \text{Tr}\{(\sigma_n^{-2} w^2 \mathbf{H}^H \mathbf{H} + v^{-1} \mathbf{I}_{N_u})^{-1}\} dv^{-1} - \lim_{v \rightarrow 0} \log v_{\hat{x}}(v) \\
&\stackrel{(f)}{=} \left[\frac{1}{N_u} \log \det \left(v^{-1} \mathbf{I}_{N_u} + \frac{w^2}{\sigma_n^2} \mathbf{H}^H \mathbf{H} \right) \right]_{v=1}^{v=0} - \lim_{v \rightarrow 0} \log \frac{1}{N_u} \text{Tr}\left\{ \left(\frac{w^2}{\sigma_n^2} \mathbf{H}^H \mathbf{H} + v^{-1} \mathbf{I}_{N_u} \right)^{-1} \right\} \\
&= \frac{1}{N_u} \log \det \left(\mathbf{I}_{N_u} + \frac{w^2}{\sigma_n^2} \mathbf{H}^H \mathbf{H} \right) \\
&\stackrel{(g)}{=} \frac{1}{N_u} \log \det \left(\mathbf{I}_{N_r} + \frac{w^2}{\sigma_n^2} \mathbf{H} \mathbf{H}^H \right) \\
&= R.
\end{aligned} \tag{47}$$

The approximation (a) is based on $\rho_i = \rho$ and $\psi_i(\rho_i) = \psi(\rho)$, equations (c) and (d) are given by (43), equation (e) comes from (42), equation (f) is based on the law $\int \text{Tr}\{s\mathbf{I} + \mathbf{A}\} ds = \log \det(s\mathbf{I} + \mathbf{A})$ and equation (g) is derived by $\det(\mathbf{I}_K + \mathbf{A}_{K \times M} \mathbf{B}_{M \times K}) = \det(\mathbf{I}_M + \mathbf{B}_{M \times K} \mathbf{A}_{K \times M})$ for any matrixes $\mathbf{A}_{K \times M}$ and $\mathbf{B}_{M \times K}$. The inequality (b) is derived by the matching condition (44)~(46) and the equality holds if and only if there exists that code whose transfer function satisfies the matching condition. In the following, we show the existence of such code whose transfer function matches the transfer function of the LMMSE estimator.

ii) Code Existence: From the “*Property of SCM Codes*”, we can see that there exist such n -layer SCM codes exist if their transfer functions satisfy (i)~(iv) and n is large enough. Therefore, in order to show the existence of such code, we only need to verify whether the matched transfer function meets the conditions (i)~(iv). It is easy to see that conditions (i) and (iv) are always satisfied by (44) and (46) respectively. From (43)~(46), we can see that $\psi(\rho)$ is continuous and differentiable in $[0, \infty)$ except at $\rho = \phi(0)$ and $\rho = \phi(1)$. Thus, Condition (iii) is satisfied. To show the monotonicity of the transfer function, we rewrite (43) by the *Matrix Inversion Lemma*.

$$\begin{aligned}
\phi(v) &= \frac{1}{N_u} \sum_{i=1}^{N_u} \left[\left\{ v - v^2 \frac{w^2}{\sigma_n^2} \mathbf{h}_i^H \left(\mathbf{I}_{N_r} + \frac{w^2 v}{\sigma_n^2} \mathbf{H} \mathbf{H}^H \right)^{-1} \mathbf{h}_i \right\}^{-1} - v^{-1} \right] \\
&= \frac{1}{N_u} \sum_{i=1}^{N_u} 1 / \left[\left(\frac{w^2}{\sigma_n^2} \mathbf{h}_i^H \left(v^{-1} \mathbf{I}_{N_r} + \frac{w^2}{\sigma_n^2} \mathbf{H} \mathbf{H}^H \right)^{-1} \mathbf{h}_i \right)^{-1} - 1 \right] \\
&= \frac{1}{N_u} \sum_{i=1}^{N_u} 1 / (f_i^{-1}(v) - 1),
\end{aligned} \tag{48}$$

where $f_i(v) = \frac{w^2}{\sigma_n^2} \mathbf{h}_i^H \left(v^{-1} \mathbf{I}_{N_r} + \frac{w^2}{\sigma_n^2} \mathbf{H} \mathbf{H}^H \right)^{-1} \mathbf{h}_i$. It is easy to verify that $f_i(v)$ is a decreasing

function with respect to v , and $\phi(v)$ is thus a decreasing function of v . With the definition of $\psi(\rho)$ based on (44)~(46), we then see that $\psi(\rho)$ is a monotonically decreasing function in $[0, \infty)$. Therefore, the matched transfer function can be constructed by the SCM code, i.e., there exists such codes that satisfy the matching conditions.

iii) Sum Rate Calculation: Thus, based on (i)~(iii), we get the achievable sum rate

$$\begin{aligned} R_{sum} &= \sum_{i=1}^{N_u} R_i \\ &\approx N_u R \\ &= \log \det \left(\mathbf{I}_{N_r} + \frac{w^2}{\sigma_n^2} \mathbf{H} \mathbf{H}^H \right), \end{aligned} \quad (49)$$

which is the exact sum capacity of the system. Therefore, we get the Theorem 1. \blacksquare

Theorem 1 shows that for a symmetric large-scale MIMO-NOMA system, the iterative detection structure is optimal, i.e., the LMMSE estimator is an optimal estimator without losing any useful information during the estimation. Although this result is derived based on the large-scale MIMO-NOMA systems, the next achievable rate analysis for the asymmetric systems shows that it also works for the general asymmetric systems.

D. Sum Capacity Achieving of Iterative LMMSE Detector for Asymmetric MIMO-NOMA Systems

For general asymmetric MIMO-NOMA system, the achievable rate region calculation of the Iterative LMMSE detection is more complicated. As we mentioned, on the one hand, all the users' transfer functions interact with each other at the estimator, i.e., the any output of the estimator relies on every variance of the input messages from the decoders. On the other hand, the transfer curve of each decoder should lies below the transfer curve of the estimator. The estimator and decoders are associated with each other. It is hard to try all the possible available transfer functions of the decoders to get the optimal code design of each user. However, the area theorem tells us that the achievable rate of every user is maximized if and only if its transfer function matches with that of the estimator if there exists some codes with that transfer function. Therefore, we can arbitrarily choose an input variances of the estimator from the decoders, and then achieve the transfer functions of the estimator. As a result, we get users' achievable rate by matching the decoders' transfer functions with the estimator.

To simplify the analysis, we add some linear constraints for the input variances of the ESE.

$$\gamma_i(v_i^{-1} - 1) = \gamma_j(v_j^{-1} - 1), \text{ for any } i, j \in \mathcal{N}_u. \quad (50)$$

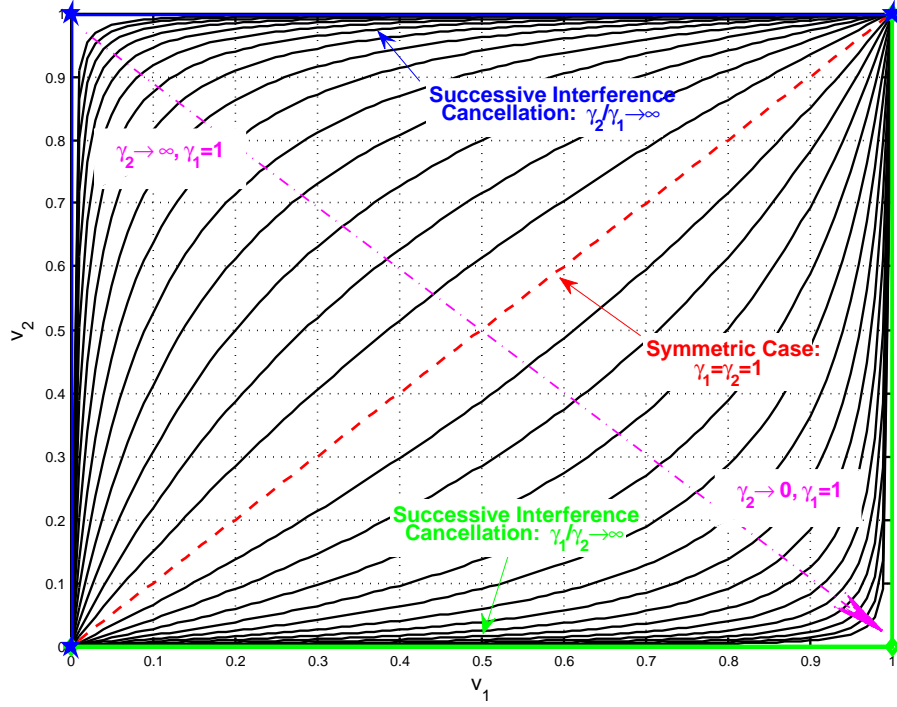


Fig. 2. Variance tracks for the different γ , where $\gamma_1 = 1$ is fixed. The variance of user i is v_i , $i = 1, 2$. When γ_2 changes from ∞ to 0, the track change from the blue curve (SIC case and the decoding order is user 1 \rightarrow user 2) to green curve (SIC case and the decoding order is user 2 \rightarrow user 1). When $\gamma_1 = \gamma_2 = 1$, it is degenerated to the symmetric case (red line).

Without loss of generality, we assume $\gamma_1 = 1$ and $\gamma_i > 0$, that is, $v_i^{-1} = 1 + \gamma_i^{-1}(v_1^{-1} - 1)$ for any $i \in \mathcal{N}_u$. Actually, the different $\gamma = [\gamma_1, \dots, \gamma_{N_u}]$ values give the different variance track during the iteration. Fig. 2 and Fig. 3 presents the variance tracks of the different γ for the two users and three users cases respectively. As we can see, when (50) concluded the symmetric case (when $w_1 = \dots = w_{N_u}$) and all the SIC points (maximal extreme points of the capacity region). If $\gamma_{k_i}/\gamma_{k_{i-1}} \rightarrow \infty$, for any $i \in \mathcal{N}_u/\{1\}$, we can get the SIC points with the decoding order $[k_1, k_2, \dots, k_{N_u}]$, which is a permutation of $[1, 2, \dots, N_u]$. The blue curve and green curves in Fig. 2 and Fig. 3 are corresponding to some of the maximal extreme points. We will also show that the different variance tracks are corresponding to the different achievable rates of the users, i.e., the user's achievable rate can be adjusted by the parameter γ .

With (50), we have

$$\mathbf{V}_{\bar{\mathbf{x}}}^{-1} = \mathbf{I}_{N_u} + \gamma_i(v_i^{-1} - 1)\mathbf{\Lambda}_{\gamma}^{-1} = \mathbf{V}_{\bar{\mathbf{x}}}^{-1}(v_i) \quad (51)$$

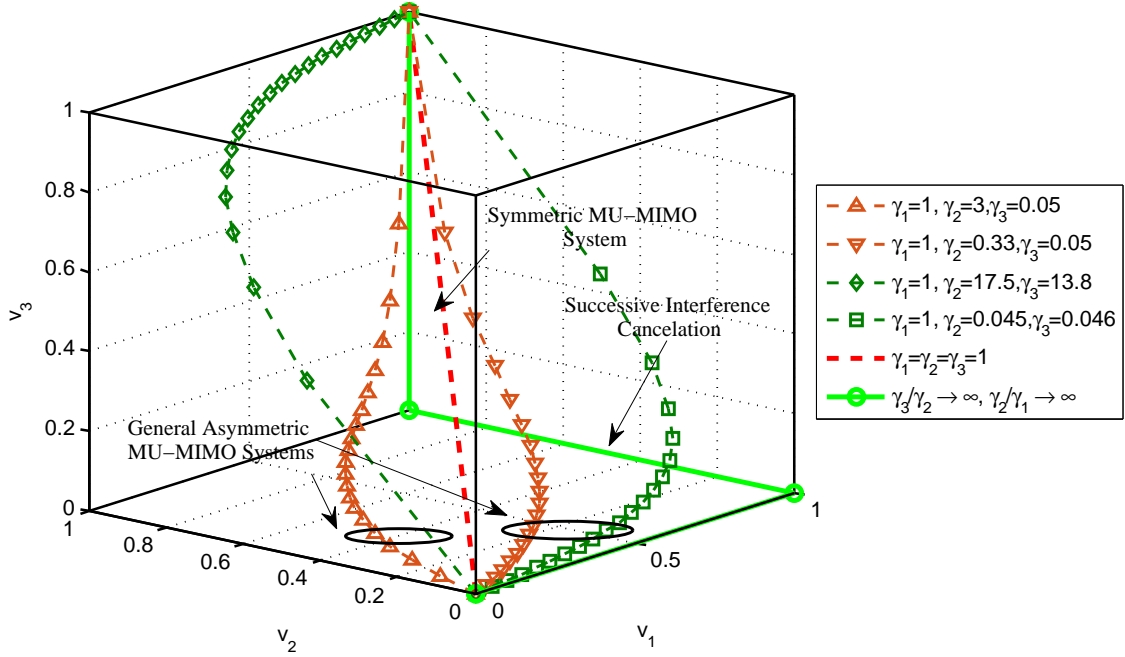


Fig. 3. Variance tracks for the different γ , where $\gamma_1 = 1$ is fixed. The variance of user i is denoted as v_i , $i = 1, 2, 3$. The variance track changes with γ_2 and γ_3 . When $\gamma_3/\gamma_2 \rightarrow \infty$ and $\gamma_2/\gamma_1 \rightarrow \infty$ (green curve), it is degenerated to a SIC case with the decoding order: user 3 \rightarrow user 2 \rightarrow user 1. When $\gamma_1 = \gamma_2 = \gamma_3 = 1$, it is degenerated to the symmetric case (red line). The other curves are the general asymmetric cases.

and

$$\begin{aligned}
 \mathbf{V}_{\hat{\mathbf{x}}} &= (\sigma_n^{-2} \mathbf{H}'^H \mathbf{H}' + \mathbf{V}_{\bar{\mathbf{x}}}^{-1})^{-1} \\
 &= (\sigma_n^{-2} \mathbf{H}'^H \mathbf{H}' + \mathbf{V}_{\bar{\mathbf{x}}}^{-1}(v_i))^{-1} \\
 &= \mathbf{V}_{\hat{\mathbf{x}}}(v_i)
 \end{aligned} \tag{52}$$

where $i \in \mathcal{N}_u$, $\Lambda_\gamma = \text{diag}(\gamma)$ is a diagonal matrix whose diagonal entries are γ . Thus, we have

$$\phi_i(\mathbf{v}_{\bar{\mathbf{x}}}) = v_{\hat{x}_i}(v_i)^{-1} - v_i^{-1} = \phi_i(v_i) = \rho_i, \tag{53}$$

For example, if we take $i = 1$, we have

$$\mathbf{V}_{\bar{\mathbf{x}}}^{-1} = \mathbf{V}_{\bar{\mathbf{x}}}^{-1}(v_1), \quad \mathbf{V}_{\hat{\mathbf{x}}} = \mathbf{V}_{\hat{\mathbf{x}}}(v_1), \quad \phi_i(\mathbf{v}_{\bar{\mathbf{x}}}) = \phi_i(v_1). \tag{54}$$

Proposition 5: Based on (53), for $i \in \mathcal{N}_u$, the matching condition (28) can be rewritten to

$$\psi_i(\rho_i) = \phi_i^{-1}(\phi_i(1)) = 1, \quad \text{for } 0 \leq \rho_i < \phi_i(1); \tag{55}$$

$$\psi_i(\rho_i) = \phi_i^{-1}(\rho_i), \quad \text{for } \phi_i(1) \leq \rho_i < \phi_i(0); \tag{56}$$

$$\psi_i(\rho_i) = 0, \quad \text{for } \phi_i(0) \leq \rho_i < \infty. \tag{57}$$

Then, we can give the users' achievable rates of the Iterative LMMSE detection for the asymmetric MIMO-NOMA systems by the following lemmas.

Lemma 1: *For the asymmetric MIMO-NOMA systems with Iterative LMMSE detection, the achievable rates of the users are*

$$R_i = \int_{v_1=1}^{v_1=0} \left[v_1 - \gamma_i^{-1} [\mathbf{V}_{\hat{\mathbf{x}}}(v_1)]_{i,i} \right] dv_1^{-1} - \log(\gamma_i), \quad (58)$$

where $i \in \mathcal{N}_u$, $\mathbf{V}_{\hat{\mathbf{x}}}(v_1) = (\sigma_n^{-2} \mathbf{H}^H \mathbf{H} + \mathbf{I}_{N_u} + (v_1^{-1} - 1) \mathbf{\Lambda}_{\gamma}^{-1})^{-1}$, and $[\cdot]_{i,i}$ denotes the i -th column and i -th row entry of the corresponding matrix.

Proof: Similarly, in this case, the achievable rate of user i can be given either by (34) at the decoder or by (35) at the estimator directly. Here, we give the first proof, and the other one is given in APPENDIX B. From (34), the achievable rate of user i is given as

$$\begin{aligned} R_i &= \int_0^\infty (\rho_i + \psi_i(\rho_i)^{-1})^{-1} d\rho_i \\ &\stackrel{(a)}{\leq} \int_{\phi_i(1)}^{\phi_i(0)} \left[\rho_i + (\phi_i^{-1}(\rho_i))^{-1} \right]^{-1} d\rho_i + \int_0^{\phi_i(1)} (1 + \rho_i)^{-1} d\rho_i \\ &\stackrel{(b)}{=} \int_{v_i=1}^{v_i=0} (v_i^{-1} + \phi_i(v_i))^{-1} d\phi_i(v_i) + \log(1 + \phi_i(v_i)) \\ &\stackrel{(c)}{=} \int_{v_i=1}^{v_i=0} v_{\hat{x}_i}(v_i) dv_{\hat{x}_i}(v_i)^{-1} - \int_{v_i=1}^{v_i=0} v_{\hat{x}_i}(v_i) dv_i^{-1} - \log v_{\hat{x}_i}(v_i = 1) \\ &\stackrel{(d)}{=} - \int_{v_1=1}^{v_1=0} \gamma_i^{-1} [\mathbf{V}_{\hat{\mathbf{x}}}(v_1)]_{i,i} dv_1^{-1} - \lim_{v_1 \rightarrow 0} \log [\mathbf{V}_{\hat{\mathbf{x}}}(v_1)]_{i,i} \\ &\stackrel{(e)}{=} \int_{v_1=1}^{v_1=0} \left[v_1 - \gamma_i^{-1} [\mathbf{V}_{\hat{\mathbf{x}}}(v_1)]_{i,i} \right] dv_1^{-1} - \log(\gamma_i). \end{aligned} \quad (59)$$

The inequality (a) is derived by (55)~(57) and the equality holds if and only if there exists that code whose transfer function satisfies the matching condition. The equations (b) ~ (d) are given by $\rho_i = \phi_i(v_i)$, (53) and (54), equation (e) comes from (51) and (52).

In APPENDIX C, similar to the Theorem 1, we show the existence of such codes whose transfer functions match the transfer functions of the LMMSE estimator. In APPENDIX D, the existence of the infinite integral of (58) is proved. ■

Lemma 2: *The achievable rate R_i of user i increases monotonously with γ_i and decreases monotonously γ_j , where $i, j \in \mathcal{N}_u$ and $j \neq i$.*

Proof: It is easy to find that $\text{mmse}_{\text{tot},i}^{\text{ese}}$ (or $\text{mmse}_{\text{tot},i}^{\text{dec}}$) increases monotonously with γ_i and decreases monotonously γ_j for $i, j \in \mathcal{N}_u$ and $j \neq i$. Thus, based on the Proposition 2, we have that R_i increases monotonously with γ_i and decreases monotonously γ_j for $j \neq i$. ■

It should be noted that although the Lemma 1 gives the achievable rate of the users and it is an integral function of channel matrix, noise variance and Λ_γ , we cannot see the specific relationship between the achievable rates and Λ_γ because of the complicated integral structure of (58). Therefore, it is very hard to show the analytical achievable rate region of the Iterative LMMSE detection. However, the sum capacity achieving of the Iterative LMMSE detection can be shown by the following theorem.

Theorem 2: *The Iterative LMMSE detection achieves the sum capacity of the MIMO-NOMA systems, i.e., $R_{\text{sum}} = \log \det (\mathbf{I}_{N_u} + \sigma_n^{-2} \mathbf{H}' \mathbf{H}^H)$.*

Proof: With (58), the achievable sum rate is

$$\begin{aligned}
R_{\text{sum}} &= \sum_{i=1}^{N_u} R_i \\
&\stackrel{(a)}{\leq} - \int_{v_1=1}^{v_1=0} \sum_{i=1}^{N_u} \left(\gamma_i^{-1} [\mathbf{V}_{\hat{\mathbf{x}}}(v_1)]_{i,i} \right) dv_1^{-1} - \lim_{v_1 \rightarrow 0} \log(v_1^{N_u} \prod_{i=1}^{N_u} \gamma_i) \\
&= - \int_{v_1=1}^{v_1=0} \text{Tr}\{\Lambda_\gamma^{-1} \mathbf{V}_{\hat{\mathbf{x}}}(v_1)\} dv_1^{-1} - \lim_{v_1 \rightarrow 0} \log(v_1^{N_u} \prod_{i=1}^{N_u} \gamma_i) \\
&\stackrel{(b)}{=} - \lim_{v_1 \rightarrow 0} \log(v_1^{N_u} \prod_{i=1}^{N_u} \gamma_i) - [\log \det ((v_1^{-1} - 1)\mathbf{I}_{N_u} + (\mathbf{I}_{N_u} + \sigma_n^{-2} \mathbf{H}'^H \mathbf{H}') \Lambda_\gamma)]_{v_1=1}^{v_1=0} \\
&= - \lim_{v_1 \rightarrow 0} \log(v_1^{N_u} \prod_{i=1}^{N_u} \gamma_i) - \lim_{v_1 \rightarrow 0} \log \det(v_1^{-1} \mathbf{I}_{N_u}) + \log \det ((\mathbf{I}_{N_u} + \sigma_n^{-2} \mathbf{H}'^H \mathbf{H}') \Lambda_\gamma) \\
&= \log \det (\mathbf{I}_{N_u} + \sigma_n^{-2} \mathbf{H}'^H \mathbf{H}'), \tag{60}
\end{aligned}$$

which is the exact system sum capacity. The inequality (a) is derived by (68) or (60), and equation (b) is based on (68) (52) and the law $\int \text{Tr}\{(s\mathbf{I} + \mathbf{A})^{-1}\} ds = \log \det(s\mathbf{I} + \mathbf{A})$. It means that the iterative detector is sum capacity-achieving for the different kinds of user-rate combinations. ■

Theorem 2 shows that for a general asymmetric MIMO-NOMA system, from the sum rate perspective, the iterative detection structure is optimal, i.e., the LMMSE estimator is an optimal estimator without losing any useful information during the estimation.

Remark 1: When $\gamma_i = 1$ for $i \in \mathcal{N}_u$, and for a symmetric system with: (i) the same rate $R_i = R$ for $i \in \mathcal{N}_u$; (ii) the same power $K_{\mathbf{x}} = w^2 I$, Theorem 2 can be degenerated to the Theorem 1. Thus, the symmetric system is a special case of the asymmetric systems.

Algorithm 1 Numeric Iterative Search Algorithm for Λ_{γ}

```

1: Input:  $\mathbf{H}$ ,  $K_{\mathbf{x}}$ ,  $\sigma_n^2$ ,  $\epsilon > 0$ ,  $N_{max}$ ,  $\mathbf{R} = [R_1, \dots, R_{N_u}]$  and calculate  $\mathbf{H}'$ .
2: If  $\mathbf{R} \in \mathcal{R}_{\mathcal{S}}$  ( $\mathcal{R}_{\mathcal{S}}$  is the capacity region given by (18))
3:   Initialize: Random choose  $\gamma = [\gamma_1, \dots, \gamma_{N_u}]$ ,  $\gamma_i > 0$  for  $\forall i \in \mathcal{N}_u$ ,
      Calculate  $\mathbf{R}^{(0)}(\gamma) = [R_1^{(0)}, \dots, R_{N_u}^{(0)}]$  by (58) and  $t = 1$ .
4:   While ( $\|\mathbf{R}^{(0)} - \mathbf{R}\|_1 > \epsilon$  or  $t < N_{max}$ )
5:     For  $i = 1 : N_u$ 
6:       fixed  $\gamma_{\sim i} = [\gamma_1, \dots, \gamma_{i-1}, \gamma_{i+1}, \dots, \gamma_{N_u}]$ ,
       find  $\gamma_i^*$  for  $R_i^{(1)}(\gamma_i = \gamma_i^*) = R_i$ , and
7:       calculate  $\mathbf{R}^{(1)}(\gamma_{\sim i}, \gamma_i^*) = [R_1^{(1)}, \dots, R_{N_u}^{(1)}]$ .
8:       While  $\|\mathbf{R}^{(1)} - \mathbf{R}\|_1 > \|\mathbf{R}^{(0)} - \mathbf{R}\|_1$ 
9:          $\gamma_i^* = (\gamma_i + \gamma_i^*)/2$  and go to step 7.
10:      End While
11:       $\gamma_i = \gamma_i^*$  and  $\mathbf{R}^{(0)} = \mathbf{R}^{(1)}$ .
12:    End For
13:     $t = t + 1$ .
14:  End While
15:  If  $t < N_{max}$ 
16:    Output:  $\gamma$ .
17:  Else
18:    go to step 21.
19:  End If
20: Else  $\mathbf{R} \notin \mathcal{R}_{\mathcal{S}}$ 
21:   Output: The given rate  $\mathbf{R}$  is outside the capacity region or not achievable.
22: End If

```

E. Practical Iterative LMMSE Detection Design for Asymmetric MIMO-NOMA Systems

Actually, the proof of the codes existence for matching conditions also gives the process of optimal codes design for the MIMO-NOMA system. It should be noted that the codes design depends on the realization of the channel matrix. Therefore, the users should know the information of the channel matrix. In addition, the codes design also depends on Λ_γ . As we cannot get a closed-form solution of the user rates with respect to Λ_γ , it is hard to find the proper Λ_γ for the given user rates. In this subsection, we propose an algorithm to search a numeric solution of Λ_γ to satisfy the rate requirement of each user.

Algorithm 1 gives a numeric iterative search of Λ_γ for the given rate \mathbf{R} , where N_{max} is the maximum iterative number, ϵ indicates the allowed precision and $\|\cdot\|_1$ denotes the 1-norm. It should be noted that γ_i^* in step 6 definitely exists and can be easily searched by dichotomy or quadratic interpolation method as R_i increases monotonously with γ_i (Lemma 2). In addition, steps 8 ~ 10 ensure that the new γ_i^* is always better than the previous one and the search program will not stop until the requirement Λ_γ is got. Experimentally, we find that the points in the system capacity region are always achievable by this numeric algorithm.

V. SOME SPECIAL CASES OF ASYMMETRIC MIMO-NOMA SYSTEMS

In Section IV, we proved that the Iterative LMMSE detection achieves the sum capacity of the asymmetric systems, but whether it achieves the whole capacity region of the asymmetric MIMO-NOMA systems is still unknown. In this section, we analyse some special cases of the asymmetric MIMO-NOMA systems. We will show that: (i) for the 2-user MIMO-NOMA system, it is proved that the Iterative LMMSE detection achieves the whole capacity region of the system, (ii) all the maximal extreme points in the capacity region of the MIMO-NOMA system can be achieved by the Iterative LMMSE detection, and (iii) for the 3-user MIMO-NOMA system, the simulation results show that the Iterative LMMSE detection can also achieve the whole capacity region of the MIMO-NOMA system.

A. Maximal Extreme Point Achieving of Iterative LMMSE Detection

As it mentioned in the *Capacity Region Domination Lemma*, the whole capacity region is dominated by a convex combination of the maximal extreme points, which has been proved that

can be achieved by the SIC method. Here, we will show that all these maximal extreme points can be achieved by the Iterative LMMSE detection when the parameter Λ_γ are properly chosen.

Corollary 1: *All the maximal extreme points of the system capacity region can be achieved by the Iterative LMMSE detection.*

Proof: For any maximal extreme point expressed in (20) with order vector $[k_1, \dots, k_{N_u}]$, we let $\gamma_{k_i}/\gamma_{k_{i-1}} \rightarrow \infty$, for any $i \in \mathcal{N}_u/\{1\}$. Therefore, similar to the green curves showed in Fig. 2 and Fig. 3, the user k_{N_u} is recovered after all the variances of other users already being zeros as $\gamma_{k_{N_u}}/\gamma_{k_{i-1}} \rightarrow \infty$, for any $i \in \mathcal{N}_u/\{1\}$.

Thus, from (58), the rate of user k_{N_u} is

$$R_{k_{N_u}} = \log \left(1 + \frac{1}{\sigma_n^2} \mathbf{h}'_{k_{N_u}} \mathbf{h}_{k_{N_u}} \right), \quad (61)$$

which is the same as that in (20). Similarly, when we recovering the k_{N_u-1} , all the users have been recovered except user k_{N_u} and user $k_{N_u} - 1$. Base on this and *Theorem 2*, we have

$$R_{k_{N_u-1}} + R_{k_{N_u}} = \log \det \left(\mathbf{I}_{|\mathcal{S}_{N_u-2}^c|} + \frac{1}{\sigma_n^2} \mathbf{H}'_{\mathcal{S}_{N_u-2}^c} \mathbf{H}_{\mathcal{S}_{N_u-2}^c} \right). \quad (62)$$

Thus, based on (61) and (62), the rate of user k_{N_u-1} is

$$R_{k_{N_u-1}} = \log \det \left(\mathbf{I}_{|\mathcal{S}_{N_u-2}^c|} + \frac{1}{\sigma_n^2} \mathbf{H}'_{\mathcal{S}_{N_u-2}^c} \mathbf{H}_{\mathcal{S}_{N_u-2}^c} \right) - \log \left(1 + \frac{1}{\sigma_n^2} \mathbf{h}'_{k_{N_u}} \mathbf{h}_{k_{N_u}} \right), \quad (63)$$

which is the same as that in (20). Continue this process and we can show all the other users' rates are the same as that of in (20). Therefore, we have *Corollary 1*. ■

These corollary shows that as the parameter Λ_γ be properly chosen, the Iterative LMMSE detection can be degenerated to the SIC methods, i.e., the SIC methods are some special cases of the proposed Iterative LMMSE detection.

B. Capacity Region Achieving for Two-user MIMO-NOMA Systems

As it is mentioned, it is very hard to calculate the specific achievable user rates from (58) for the general asymmetric MIMO-NOMA systems. In this subsection, we show that the Iterative LMMSE detection can achieve the whole capacity region of two-user MIMO-NOMA systems.

Theorem 3: *The Iterative LMMSE detection achieves the whole capacity region of two-user MIMO-NOMA systems as follows.*

$$\begin{cases} R_1 \leq \log(1 + \frac{1}{\sigma_n^2} \mathbf{h}'_1 \mathbf{h}_1), \\ R_2 \leq \log(1 + \frac{1}{\sigma_n^2} \mathbf{h}'_2 \mathbf{h}_2), \\ R_1 + R_2 \leq \log \det (\mathbf{I}_2 + \sigma_n^{-2} \mathbf{H}'^H \mathbf{H}). \end{cases} \quad (64)$$

Proof: The pentagon in Fig. 4 indicates the capacity region of two-user MIMO-NOMA system, which is dominated by segment AB and point A and point B are two maximal extreme points. Without loss of generality, we let $\gamma_1 = 1$ and $\gamma_2 = \gamma \in [0, \infty)$. From *Theorem 2*, we get

$$R_{sum} = R_1 + R_2 = \log \det (\mathbf{I}_2 + \sigma_n^{-2} \mathbf{H}'^H \mathbf{H}'), \quad (65)$$

which is the exact sum capacity of the system.

In addition, as we discussed in *Corollary 1*, when γ changes from 0 to ∞ , R_1 reduces from $\log \left(1 + \frac{1}{\sigma_n^2} \mathbf{h}'_1^H \mathbf{h}'_1 \right)$ to $\log \det (\mathbf{I}_2 + \sigma_n^{-2} \mathbf{H}'^H \mathbf{H}') - \log \left(1 + \frac{1}{\sigma_n^2} \mathbf{h}'_1^H \mathbf{h}'_1 \right)$, and R_2 increases from $\log \det (\mathbf{I}_2 + \sigma_n^{-2} \mathbf{H}'^H \mathbf{H}') - \log \left(1 + \frac{1}{\sigma_n^2} \mathbf{h}'_2^H \mathbf{h}'_2 \right)$ to $\log \left(1 + \frac{1}{\sigma_n^2} \mathbf{h}'_2^H \mathbf{h}'_2 \right)$. As the R_1 and R_2 are both continuous functions of γ , from (65), we can see that when the parameter γ changes from 0 to ∞ , the point (R_1, R_2) moves from maximal extreme point B to maximal extreme point A along the segment AB. It means that the Iterative LMMSE detection can achieve any point on the segment AB. Therefore, the Iterative LMMSE detection achieves the whole capacity region, because it is dominated by the segment AB. ■

Actually, for the simple two-user case, we can give the specific expressions of R_1 and R_2 . The following corollary is derived directly based on *Lemma 1*.

Corollary 2: For two-user MIMO-NOMA system with Iterative LMMSE detection, the user rates are given by

$$\begin{cases} R_1 = \frac{1}{2} \log(\gamma \det(\mathbf{A})) + \frac{a_{22}\gamma - a_{11}}{2\eta} \log \frac{a_{22}\gamma + a_{11} - \eta}{a_{22}\gamma + a_{11} + \eta}, \\ R_2 = \frac{1}{2} \log(\gamma^{-1} \det(\mathbf{A})) - \frac{a_{22}\gamma - a_{11}}{2\eta} \log \frac{a_{22}\gamma + a_{11} + \eta}{a_{22}\gamma + a_{11} - \eta}, \end{cases} \quad (66)$$

where $\mathbf{A} = \sigma_n^{-2} \mathbf{H}'^H \mathbf{H}' + \mathbf{I}_2 = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ and $\eta = \sqrt{a_{22}^2 \gamma^2 + 2(2a_{21}a_{12} - a_{22}a_{11})\gamma + a_{11}^2}$. It is easy to find that η is a real number as \mathbf{A} is positive definite and $\gamma \geq 0$.

Remark 2: It should be noted from (66) that R_1 and R_2 are not linear functions of γ . It is easy to check that $R_1 + R_2 = \log \det (\mathbf{I}_2 + \sigma_n^{-2} \mathbf{H}'^H \mathbf{H}')$, and when $\gamma \rightarrow 0$ (or $\gamma \rightarrow \infty$), the limit of (R_1, R_2) in (66) converges to the maximal point B (or A) in Fig. 4. When the parameter γ changes from 0 to ∞ , the point (R_1, R_2) can achieve any point on the segment AB in Fig. 4. It also shows another proof of *Theorem 3*. In addition, the achievable rates of TDMA and OFDMA are strictly smaller than that of the Iterative LMMSE NOMA systems. Fig. 5 and Fig. 6 present the rate curves of R_1 and R_2 respect to the parameter γ . It verifies that R_2 increases monotonously with the γ and $R_1 + R_2$ always equals to the system sum capacity.

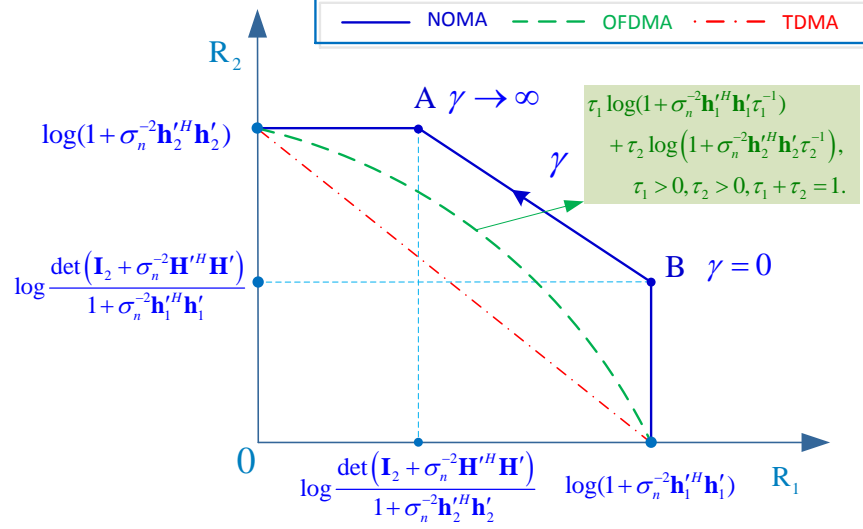


Fig. 4. Capacity region achieving of Iterative LMMSE detector for two-user MIMO-NOMA system. When the parameter γ changes from 0 to ∞ , point (R_1, R_2) moves from maximal extreme point B to maximal extreme point A along segment AB.

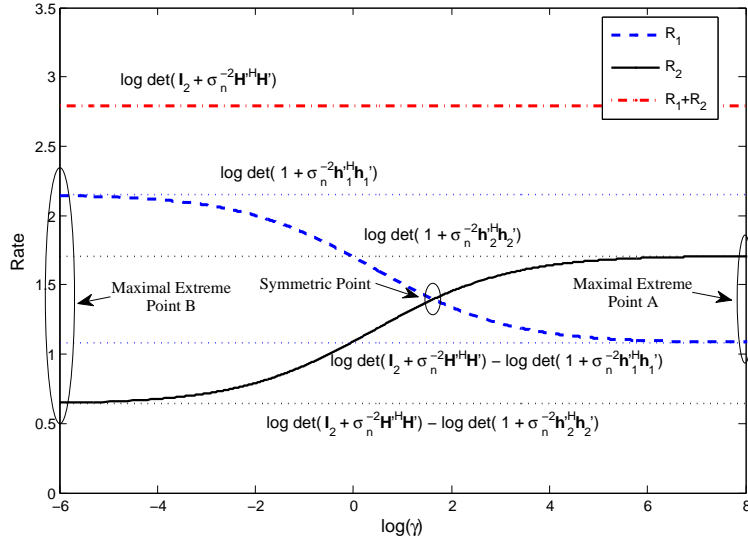


Fig. 5. Relationship between the user rates and parameter γ of the Iterative LMMSE detection for two-user MIMO-NOMA system. $N_r = 2$, $N_u = 2$, $\sigma_N^2 = 0.5$ and $\mathbf{H} = [1.32 \ -1.31; -1.43 \ 0.74]$.

C. Capacity Region Achieving for Three-user MIMO-NOMA Systems

For the three-user MIMO-NOMA systems, it is hard to get a closed-form solution of the user rates. Therefore, it is difficult to show that the Iterative LMMSE detection can achieve the whole capacity region. However, the user rates can be solved by numerical calculation for (58). Fig. 7 shows the relationships between the user rates and γ_1 and γ_2 , where $N_r = 2$, $N_u = 3$, $\sigma_N^2 = 0.5$,

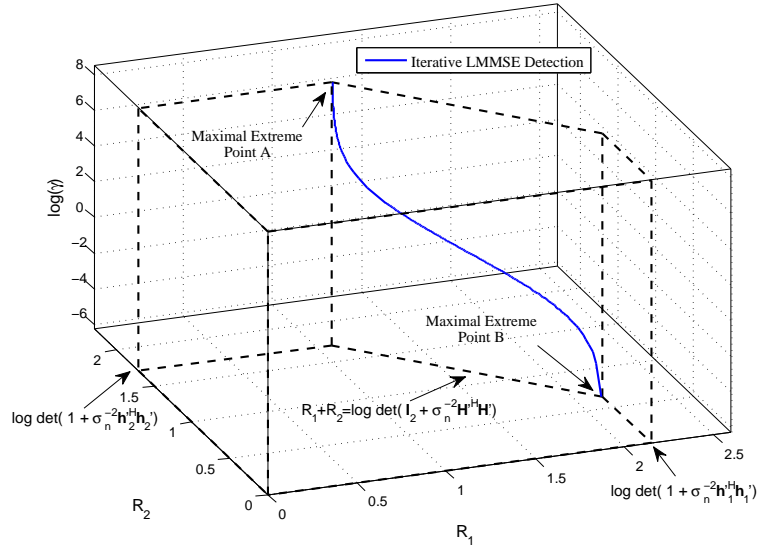


Fig. 6. Relationship between the user rates and parameter γ of the Iterative LMMSE detection for two-user MIMO-NOMA system. $N_r = 2$, $N_u = 2$, $\sigma_N^2 = 0.5$ and $\mathbf{H} = [1.32 \ -1.31; -1.43 \ 0.74]$.

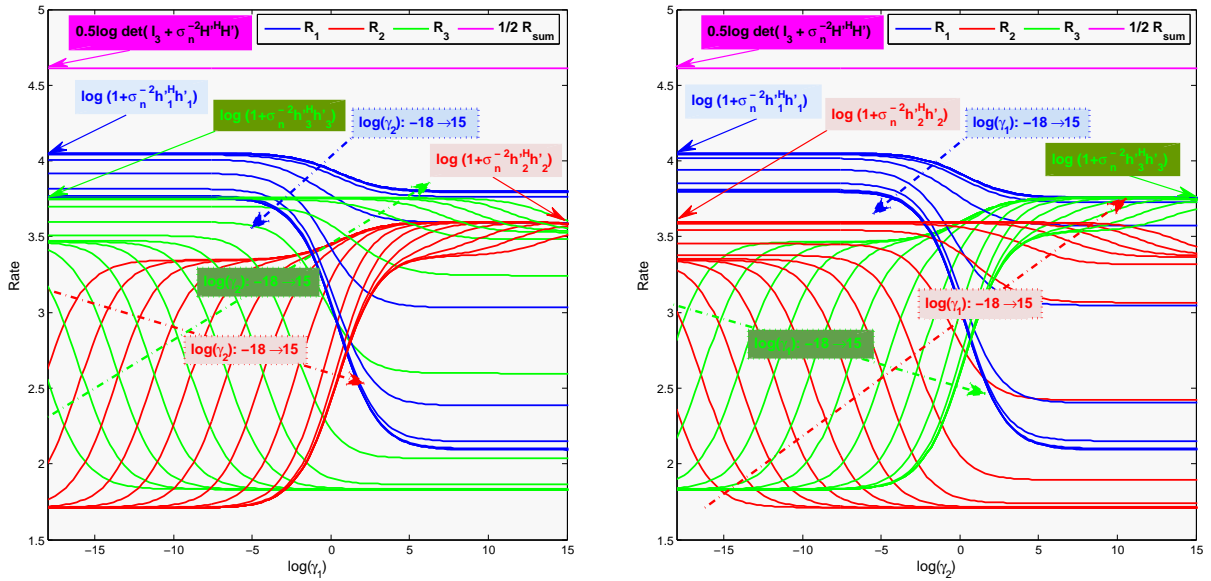


Fig. 7. Relationship between the user rates and parameters (γ_1, γ_2) of the Iterative LMMSE detection for three-user MIMO system. $N_r = 2$, $N_u = 3$, $\sigma_N^2 = 0.5$ and $\mathbf{H} = [0.678 \ 0.603 \ 0.655; 0.557 \ 0.392 \ 0.171]$.

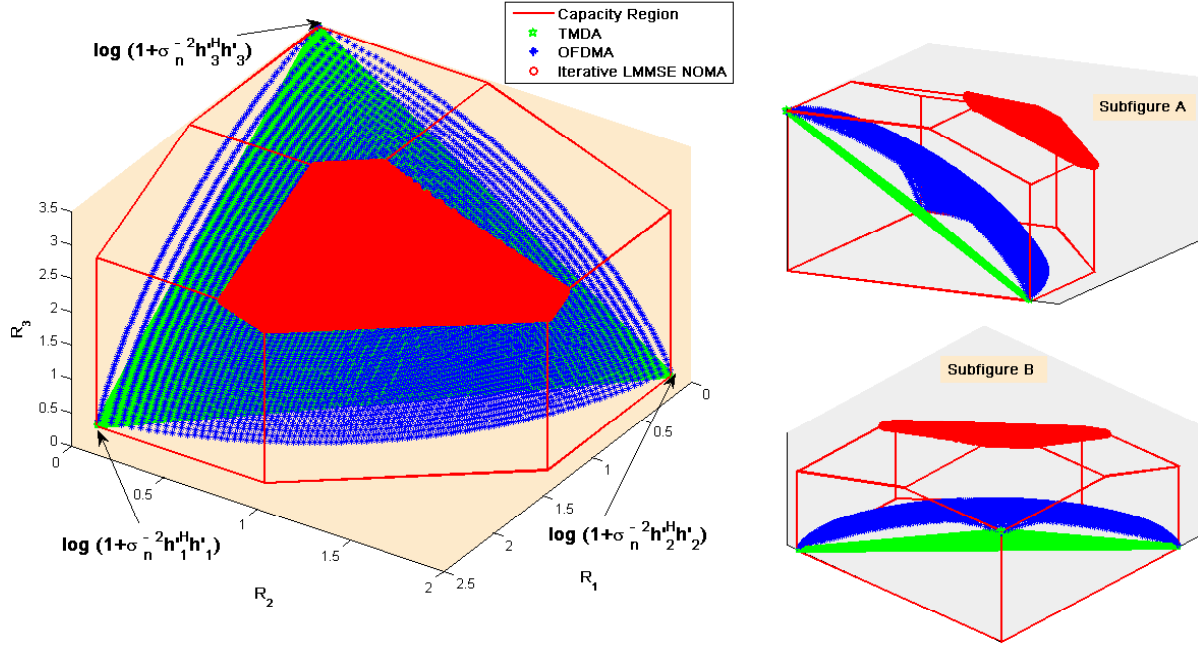


Fig. 8. Achievable user rates for all (γ_1, γ_2) of the Iterative LMMSE detection for three-user MIMO-NOMA system. $N_r = 3$, $N_u = 3$, $\sigma_N^2 = 0.5$ and $\mathbf{H} = [1.95 \ 1.28 \ -2.53; -0.31 \ -0.16 \ 2.22; 0.55 \ 1.08 \ -1.98]$. Subfigure A and Subfigure B are the same figure with different rotated viewpoints.

and $\mathbf{H} = [0.678 \ 0.603 \ 0.655; 0.557 \ 0.392 \ 0.171]$. It should be noted that although the user rates change with γ_1 and γ_2 , the sum rate $R_{sum} = \sum_{i=1}^3 R_i$ is constant and equals to the system sum capacity. In Fig. 7, we can also see that the user rate R_2 increases monotonously with γ_1 , but R_1 and R_3 decrease monotonously with γ_1 . Similarly, the user rate R_3 increases monotonously with γ_2 , but R_1 and R_2 decrease monotonously with γ_2 . In Fig. 8, the system capacity region is the polygonal consisted by the red lines, which is dominated by the red hexagonal face. The red points in Fig. 8 are the achievable points of the Iterative LMMSE detection. It shows that as we change the values of γ_1 and γ_2 , the achievable points of the Iterative LMMSE detection can reach any point on the dominate hexagonal face. Therefore, for the three-user MIMO-NOMA systems the Iterative LMMSE detection can also achieve the whole system capacity region, i.e., the Iterative LMMSE detection is an optimal detection. In addition, we can see that the achievable rates of TDMA and OFDMA are strictly smaller than that of the Iterative LMMSE NOMA systems. It should be noted that the results in this paper can also apply to the case (like Fig. 7) that the number of users is larger than the number of antennas, i.e., $N_u > N_r$.

VI. CONCLUSION

We studied an Iterative LMMSE detector for MIMO-NOMA systems, which has a low-complexity as the distributed calculations replace the overall processing. The achievable rate region of the Iterative LMMSE detector has been analysed for the symmetric and asymmetric MIMO-NOMA systems. For the symmetric MIMO-NOMA systems, it is proved that the Iterative LMMSE detector is capacity achieving, and for the asymmetric MIMO-NOMA systems, we prove that the Iterative LMMSE detector is sum capacity achieving. In addition, it is showed that with the carefully designed Iterative LMMSE detector, all the maximal extreme points in the capacity region of asymmetric MIMO-NOMA systems are achievable, and the whole capacity regions of two-user and three-user asymmetric systems are also achievable.

APPENDIX A

AN ALTERNATIVE PROOF OF (47)

An alternative proof of (47) can be derived based on (35) at the estimator directly as follows.

$$\begin{aligned}
R_i &\stackrel{(a)}{\leq} \int_{\phi_i(\mathbf{v}_{\bar{\mathbf{x}}})=0}^{\phi_i(\mathbf{v}_{\bar{\mathbf{x}}})=\infty} v_{\hat{x}_i}(\mathbf{v}_{\bar{\mathbf{x}}}) d\phi_i(\mathbf{v}_{\bar{\mathbf{x}}}) \\
&\stackrel{(b)}{\approx} \int_{\phi(v)=0}^{\phi(v)=\infty} v_{\hat{x}}(v) d\phi(v) \\
&\stackrel{(c)}{=} \int_{\phi(v)=0}^{\phi(v)=\phi(1)} (1 + \phi(v))^{-1} d\phi(v) + \int_{\phi(v)=\phi(1)}^{\phi(v)=\phi(0)} v_{\hat{x}}(v) d\phi(v) \\
&= -\log v_{\hat{x}}(v=1) + \int_{v=1}^{v=0} v_{\hat{x}}(v) d(v_{\hat{x}}(v)^{-1} - v^{-1}) \\
&= \int_{v=1}^{v=0} v_{\hat{x}}(v) dv_{\hat{x}}(v)^{-1} - \int_{v=1}^{v=0} v_{\hat{x}}(v) dv^{-1} - \log v_{\hat{x}}(v=1) \\
&= \frac{1}{N_u} \log \det \left(\mathbf{I}_{N_r} + \frac{w^2}{\sigma_n^2} \mathbf{H} \mathbf{H}^H \right) \\
&= R.
\end{aligned} \tag{67}$$

The inequality (a) is based on the area property (35) and the equality holds if and only if there exists that code whose transfer function satisfies the matching condition. The approximation (b)

come from (42). Equation (c) is based on the fact that the value region $[0, \phi(1)]$ of $\phi(v)$ is corresponding to a single value $v_i = 1$, and $v_{\hat{x}}(v) = 0$ if $\phi(v) > \phi(0)$. The following equations in (67) are similar with that of the (47).

APPENDIX B

AN ALTERNATIVE PROOF OF (59)

The achievable rate of user i can also be derived based on (35) at the estimator directly as follows.

$$\begin{aligned}
R_i &\stackrel{(a)}{\leq} \int_{\phi_i(\mathbf{v}_{\bar{\mathbf{x}}})=0}^{\phi_i(\mathbf{v}_{\bar{\mathbf{x}}})=\infty} v_{\hat{x}_i}(\mathbf{v}_{\bar{\mathbf{x}}}) d\phi_i(\mathbf{v}_{\bar{\mathbf{x}}}) \\
&\stackrel{(b)}{=} \int_{\phi_i(v_i=1)}^{\phi_i(v_i=0)} v_{\hat{x}_i}(v_i) d\phi_i(v_i) + \int_0^{\phi_i(v_i=1)} (1 + \phi_i(v_i))^{-1} d\phi_i(v_i) \\
&\stackrel{(c)}{=} \int_{v_i=1}^{v_i=0} v_{\hat{x}_i}(v_i) d(v_{\hat{x}_i}(v_i)^{-1} - v_i^{-1}) - \log v_{\hat{x}_i}(v_i = 1) \\
&\stackrel{(d)}{=} - \int_{v_1=1}^{v_1=0} \gamma_i^{-1} [\mathbf{V}_{\hat{\mathbf{x}}}(v_1)]_{i,i} dv_1^{-1} - \lim_{v_1 \rightarrow 0} \log(\gamma_i v_1), \tag{68}
\end{aligned}$$

Similarly, the inequality (a) is derived by the matching condition (55)~(57) and the equality holds if and only if there exists that code whose transfer function satisfies the matching condition. Equations (b) and (c) are based on (52)~(54).

APPENDIX C

THE CODES EXISTENCE OF LEMMA 1

From the “*Property of SCM Codes*”, we can see that there exist such n -layer SCM codes whose transfer function satisfies (i)~(iv) when n is large enough. Therefore, it only needs to check the matched transfer function meets the conditions (i)~(iv) in order to show the existence of such code. It is easy to see that conditions (i) and (iv) are always satisfied by (55) and (56) respectively. From (53)~(57), we can see that $\psi_i(\rho_i)$ is continuous and differentiable in $[0, \infty)$ except at $\rho_i = \phi_i(0)$ and $\rho_i = \phi_i(1)$. Thus, Condition (iii) is satisfied. To show the monotonicity

of the transfer function, we first rewritten (43) by the random matrix theorem as

$$\begin{aligned}
\phi_i(v_i) &= \left[v_i - v_i^2 \frac{w^2}{\sigma_n^2} \mathbf{h}_i^H \left(\mathbf{I}_{N_r} + \frac{w^2 v_i}{\sigma_n^2} \mathbf{H} \mathbf{H}^H \right)^{-1} \mathbf{h}_i \right]^{-1} - v_i^{-1} \\
&= \left[\left(\frac{w^2}{\sigma_n^2} \mathbf{h}_i^H \left(v_i^{-1} \mathbf{I}_{N_r} + \frac{w^2}{\sigma_n^2} \mathbf{H} \mathbf{H}^H \right)^{-1} \mathbf{h}_i \right)^{-1} - 1 \right]^{-1} \\
&= 1 / (f_i^{-1}(v_i) - 1),
\end{aligned} \tag{69}$$

where $f_i(v_i) = \frac{w^2}{\sigma_n^2} \mathbf{h}_i^H \left(v_i^{-1} \mathbf{I}_{N_r} + \frac{w^2}{\sigma_n^2} \mathbf{H} \mathbf{H}^H \right)^{-1} \mathbf{h}_i$. It is easy to check that $f_i(v_i)$ is a decreasing function with respect to v_i , and $\phi_i(v_i)$ is thus a decreasing function of v . With the definition of $\psi(\rho)$ from (55)~(57), we then see that $\psi_i(\rho_i)$ is a monotonically decreasing function in $[0, \infty)$. Therefore, the matched transfer function can be constructed by the SCM code, i.e., there exists such codes that satisfy the matching condition.

APPENDIX D

THE EXISTENCE OF INFINITE INTEGRAL (58)

With (58), we have

$$\begin{aligned}
R_i &= - \int_{v_1=1}^{v_1=0} \gamma_i^{-1} [\mathbf{V}_{\hat{\mathbf{x}}}(v_1)]_{i,i} dv_1^{-1} - \lim_{v_1 \rightarrow 0} \log(\gamma_i v_1) \\
&= - \int_0^\infty [(\mathbf{A}_\gamma + s \mathbf{I}_{N_u})^{-1}]_{i,i} ds - \lim_{s \rightarrow \infty} \log(\gamma_i s^{-1}), \quad s = v_1^{-1}, \mathbf{A}_\gamma = \mathbf{\Lambda}_\gamma^{1/2} \left(\sigma_n^{-2} \mathbf{H}'^H \mathbf{H}' + \mathbf{I}_{N_u} \right) \mathbf{\Lambda}_\gamma^{1/2} \\
&= - \int_0^\infty \mathbf{u}_i^H (\mathbf{\Lambda}_{A_\gamma} + s \mathbf{I}_{N_u})^{-1} \mathbf{u}_i ds - \lim_{s \rightarrow \infty} \log(\gamma_i s^{-1}), \quad \mathbf{A}_\gamma = \mathbf{U}^H \mathbf{\Lambda}_{A_\gamma} \mathbf{U}, \mathbf{u}_i \rightarrow i\text{th column of } \mathbf{U} \\
&= - \int_0^\infty \sum_{j=1}^{N_u} \|u_{ij}\|^2 (\lambda_{\mathbf{A}_\gamma, j} + s)^{-1} ds - \lim_{s \rightarrow \infty} \log(\gamma_i s^{-1}), \quad \lambda_{\mathbf{A}_\gamma, j} \rightarrow i\text{th diagonal element of } \mathbf{\Lambda}_{A_\gamma} \\
&= \sum_{j=1}^{N_u} \|u_{ij}\|^2 \log(\lambda_{\mathbf{A}_\gamma, j}) - \log(\gamma_i)
\end{aligned} \tag{70}$$

Thus, we show the existence of the infinite integral (59) or (68), i.e., R_i has a finite value.

REFERENCES

- [1] METIS, “Proposed solutions for new radio access,” *Mobile and wireless communications enablers for the 2020 information society (METIS)*, Deliverable D.2.4, Feb. 2015.
- [2] “5G radio access: requirements, concepts and technologies,” *NTT DOCOMO, Inc., Tokyo, Japan, 5G Whitepaper*, Jul. 2014.
- [3] B. Kim and W. Chung, “Uplink NOMA with Multi-Antenna,” in *Proc. of IEEE VTC 2015-Spring, Scotland, UK*, 2015.
- [4] S. Chen, K. Peng and H. Jin, “A suboptimal scheme for uplink NOMA in 5G systems,” in *Proc. of IEEE International Wireless Communications and Mobile Computing Conference (IWCMC)*, Aug. 2015.
- [5] M. Al-Imari, P. Xiao, M. A. Imran, and R. Tafazolli, “Uplink non-orthogonal multiple access for 5g wireless networks,” in *Proc. of the 11th Int. Symp. on Wireless Commun. Systems (ISWCS), Barcelona, Spain*, Aug 2014, pp. 781-785.
- [6] Y. Saito, Y. Kishiyama, A. Benjebbour, T. Nakamura, A. Li, and K. Higuchi, “Non-orthogonal multiple access (NOMA) for cellular future radio access,” in *Proc. IEEE Vehicular Technology Conference, Dresden, Germany*, Jun. 2013.
- [7] Z. Ding, Z. Yang, P. Fan, and H. V. Poor, “On the performance of non-orthogonal multiple access in 5G systems with randomly deployed users,” *IEEE Signal Process. Letters*, vol. 21, no. 12, pp. 1501-1505, Dec 2014.
- [8] Z. Ding, M. Peng, and H. V. Poor, “Cooperative non-orthogonal multiple access in 5G systems,” *IEEE Commun. Lett.*, vol. 19, no. 8, pp. 1462- 1465, Aug. 2015.
- [9] Z. Yang, Z. Ding, P. Fan, and G. K. Karagiannidis, “On the Performance of Non-orthogonal Multiple Access Systems With Partial Channel Information,” *IEEE Trans. Commun.*, vol. 64, no. 2, pp. 654- 667, Feb. 2015.
- [10] Z. Ding, F. Adachi, and H. V. Poor, “The application of MIMO to non-orthogonal multiple access,” *IEEE Trans. Wireless Commun.*, 2016, submitted for publication.
- [11] D. Argas, D. Gozalvez, D. Gomez-Barquero, and N. Cardona, “MIMO for DVB-NGH, the next generation mobile TV broadcasting,” *IEEE Commun. Mag.*, vol. 51, no. 7, pp. 130-137, Jul. 2013.
- [12] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, “Scaling up MIMO: Opportunities and challenges with very large arrays,” *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40-60, Jan. 2013.
- [13] E. Biglieri, R. Calderbank, A. Constantinides, A. Goldsmith, A. Paulraj, and H. V. Poor, *MIMO Wireless Communications*. Cambridge University Press, Cambridge, 2007.
- [14] T. L. Marzetta, “Noncooperative cellular wireless with unlimited numbers of base station antennas,” *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590-3600, Nov. 2010.
- [15] D. Micciancio, “The hardness of the closest vector problem with preprocessing,” *IEEE Transactions on Information Theory*, vol. 47, no. 3, pp. 1212-1215, Mar. 2001.
- [16] S. Verdú, “Optimum multi-user signal detection,” Ph.D. dissertation, Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, Aug. 1984.
- [17] S. Verdú and H. V. Poor, “Backward, forward and backward-forward dynamic programming models under commutativity conditions,” in *Proc. the 23rd IEEE Conf. on Decision and Control (CDC’84)*, Las Vegas, NV, Dec. 1984, pp. 1081-1086.
- [18] S. Verdú, “Optimum sequence detection of asynchronous multipleaccess communications,” in *Abstr. IEEE International Symposium on Information Theory (ISIT’83)*, St. Jovite, Canada, Sep. 1983, p. 80.
- [19] S. Verdú and H. V. Poor, “Abstract dynamic programming models under commutativity conditions,” *SIAM Journal on Control and Optimization*, vol. 25, no. 4, pp. 990-1006, Jul. 1987.
- [20] Tse David and Pramod Viswanath, *Fundamentals of wireless communication*. Cambridge university press, 2005.
- [21] H. A. Loeliger, “An introduction to factor graphs,” *IEEE Signal Processing Mag.*, pp. 28-41, Jan. 2004.

- [22] H. A. Loeliger, "Least squares and Kalman filtering on Forney graphs," in *Codes, Graphs, and Systems, (festschrift in honor of David Forney on the occasion of his 60th birthday)*, R.E. Blahut and R. Koetter, eds., Kluwer, 2002, pp. 113-135.
- [23] H. A. Loeliger, J. Hu, S. Korl, Q. Guo and L. Ping, "Gaussian message passing on linear models: an update," *Int. Symp. on Turbo codes and Related Topics*, Apr. 2006.
- [24] O. Axelsson, *Iterative Solution Methods*. Cambridge, UK: Cambridge University Press, 1994.
- [25] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Calculation. Numerical Methods*. Prentice Hall, 1989.
- [26] X. Gao, L. Dai, C. Yuen, and Y. Zhang, "Low-Complexity MMSE Signal Detection Based on Richardson Method for Large-Scale MIMO Systems," in *IEEE 80th Vehicular Technology Conference (VTC Fall)*, Sept. 2014, pp. 1-5.
- [27] Lei Liu, Yuen Chau, Yong Liang Guan, Ying Li and Yuping Su, "A Low-Complexity Gaussian Message Passing Iterative Detection for Massive MU-MIMO Systems," in *Proc IEEE International Conference on Information, Communications and Signal Processing (ICICS)*, Singapore, Dec. 2015.
- [28] A. Montanari, B. Prabhakar, and David Tse, "Belief Propagation Based Multi-User Detection," *Proceedings*, Vol. 43, 2005.
- [29] P. Rusmevichientong and B. Van Roy, "An analysis of belief propagation on the turbo decoding graph with Gaussian densities", *IEEE Trans. Inform. Theory*, vol. 47, pp.745-765, 2001.
- [30] S. Yoon and C. Chae, "Low-Complexity MIMO Detection Based on Belief Propagation Over Pairwise Graphs," *IEEE Trans. on TVT*, vol. 63, no. 5, pp. 2363-2377, 2014.
- [31] T. M. Cover and J. A. Thomas, *Elements of Information Theory-Second Edition*. New York: Wiley, 2006.
- [32] A. E. Gamal and Young-Han Kim. *Network information theory*. Cambridge University Press, January 2012.
- [33] S. Verdú, *Multiuser Detection*. Cambridge, UK: Cambridge University Press, 1998.
- [34] G. D. Golden, G. J. Foschini, R. A. Valenzuela, and P. W. Wolniansky, "Detection algorithm and initial laboratory results using V-BLAST spacetime communication architecture," *Electron. Lett.*, vol. 35, no. 1, pp. 14-16, January 1999.
- [35] X. Wang and H. Poor, "Iterative (turbo) soft interference cancellation and decoding for coded CDMA," *IEEE Trans. Commun.*, vol. 47, no. 7, pp. 1046-1061, Jul. 1999.
- [36] W. J. Choi, K. W. Cheong, and J. M. Cioffi, "Iterative soft interference cancellation for multiple antenna systems," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, 2000, vol. 1, pp. 304-309.
- [37] C. Studer, S. Fateh, and D. Seethaler, "ASIC implementation of soft-input soft-output MIMO detection using MMSE parallel interference cancellation," *IEEE J. Solid-State Circuits*, vol. 46, no. 7, pp. 1754-1765, Jul. 2011.
- [38] P. Li, L. Liu, K. Y. Wu, and W. K. Leung, "Interleave-division multiple-access (IDMA) communications," in *Proc. Int. Symp. Turbo Codes Related Topics, Brest, France, Sept. 2003*, pp. 173-180.
- [39] P. Wang, J. Xiao, and P. Li, "Comparison of Orthogonal and Non-Orthogonal Approaches to Future Wireless Cellular Systems," *IEEE Vehicular Technology Magazine*, vol. 1, no. 3, pp. 4-11, Sept. 2006.
- [40] L. Ping, L. Liu, K. Y. Wu, and W. K. Leung, "Approaching the Capacity of Multiple Access Channels Using Interleaved Low-Rate Codes," *IEEE Commun. letters*, vol. 8, no. 1, pp. 4-6, Jan. 2004.
- [41] Q. Guo and L. Ping, "LMMSE turbo equalization based on factor graphs," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 2, pp. 311-319, 2008.
- [42] S. Wu, L. Kuang, Z. Ni, J. Lu, D. DavidHuang, and Q. Guo, "Low-Complexity Iterative Detection for Large-Scale Multiuser MIMO-OFDM Systems Using Approximate Message Passing," *IEEE Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 902-915, 2014.
- [43] A. Sanderovich, M. Peleg, and S. Shamai, "LDPC coded MIMO multiple access with iterative joint decoding," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1437-1450, Apr. 2005.

- [44] G. Caire, R. Muller, and T. Tanaka, "Iterative multiuser joint decoding: Optimal power allocation and low-complexity implementation," *IEEE Trans. Inf. Theory*, vol. 50, no. 9, pp. 1950-1973, Sep. 2004.
- [45] X. Yuan, L. Ping, C. Xu and A. Kavcic, "Achievable Rates of MIMO Systems With Linear Precoding and Iterative LMMSE Detection," *IEEE Trans. Inf. Theory*, vol. 60, no.11, pp. 7073-7089, Oct. 2014.
- [46] K. Bhattad and K. R. Narayanan, "An MSE-based transfer chart for analyzing iterative decoding schemes using a Gaussian approximation," *IEEE Trans. Inf. Theory*, vol. 53, no. 1, pp. 22-38, Jan. 2007.
- [47] D. Guo, S. Shamai, and S. Verdú, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1261-1282, Apr. 2005.
- [48] A. Ashikhmin, G. Kramer, and S. ten Brink, "Extrinsic information transfer functions: Model and erasure channel properties," *IEEE Trans. Inf. Theory*, vol. 50, no. 11, pp. 2657-2673, Nov. 2004.
- [49] S. ten Brink, "Convergence behavior of iteratively decoded parallel concatenated codes," *IEEE Trans. Commun.*, vol. 49, no. 10, pp. 1727-1737, Oct. 2001.
- [50] K. S. Andrews, D. Divsalar, S. Dolinar, J. Hamkins, C. R. Jones, F. Pollara, "The Development of Turbo and LDPC Codes for Deep Space Applications," *Proceedings of the IEEE*, Vol. 95, No. 11, Nov. 2007.
- [51] T. J. Richardson and R. L. Urbanke, "The capacity of low-density paritycheck codes under message-passing decoding," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 599-618, Feb. 2001.
- [52] T. S. Han, "The capacity region of general multiple-access channel with certain correlated sources," *Inf. Control*, vol. 40, no. 1, pp. 37-60, 1979.
- [53] S. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ, USA: Prentice-Hall, 1993.
- [54] H. V. Poor and S. Verdú, "Probability of error in MMSE multiuser detection," *IEEE Trans. Inf. Theory*, vol. IT-43, no. 3, pp. 835-847, May 1997.
- [55] D. Guo, Y. Wu, S. Shamai, and S. Verdú, "Estimation in Gaussian noise: Properties of the minimum mean-square error," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 2371-2385, Apr. 2011.
- [56] U. Wachsmann, R. F. H. Fischer, and J. B. Huber, "Multilevel codes: Theoretical concepts and practical design rules," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1361-1391, Jul. 1999.
- [57] S. Gadkari and K. Rose, "Time-division versus superposition coded modulation schemes for unequal error protection," *IEEE Trans. Commun.*, vol. 47, no. 3, pp. 370-379, Mar. 1999.